



The Urge to Detect, the Need to Clarify

Gricean Perspectives on Information, Misinformation and Disinformation

Søe, Sille Obelitz

Publication date:
2016

Document license:
[CC BY-NC-ND](#)

Citation for published version (APA):
Søe, S. O. (2016). *The Urge to Detect, the Need to Clarify: Gricean Perspectives on Information, Misinformation and Disinformation*. Det Humanistiske Fakultet, Københavns Universitet.

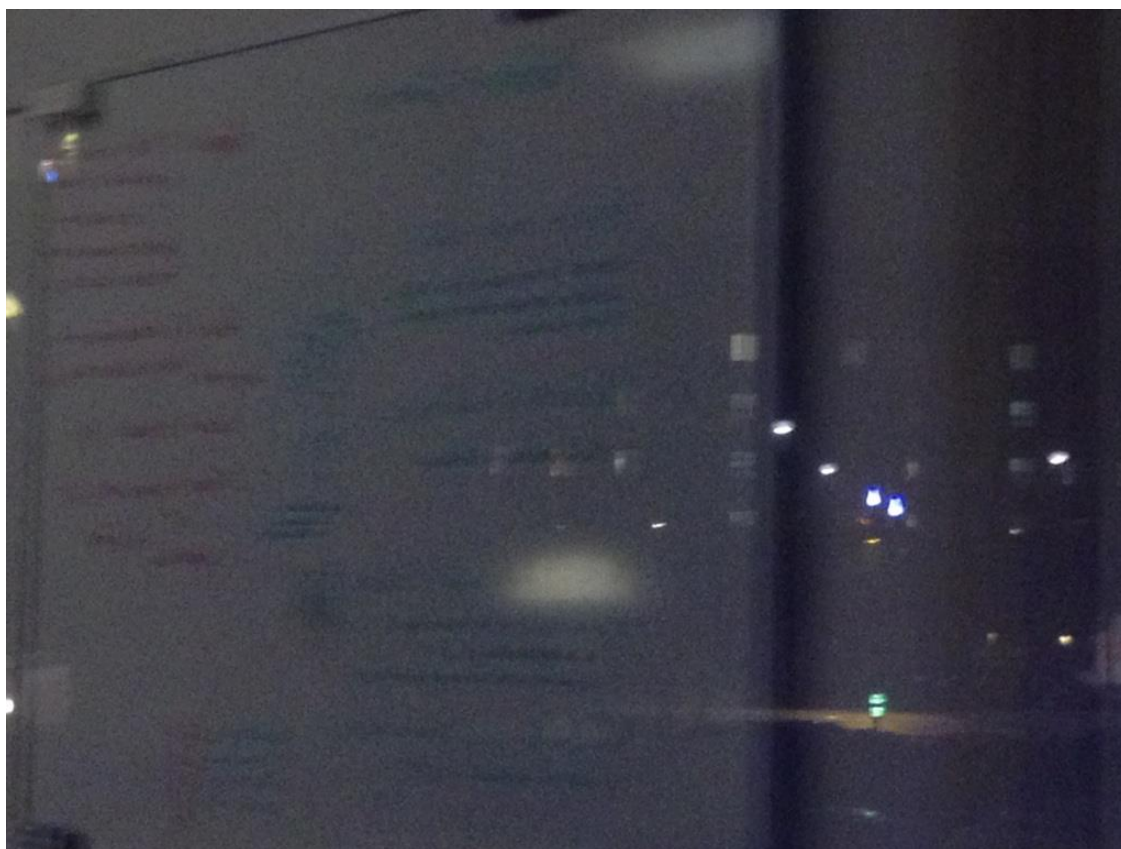


PhD Thesis

Sille Obelitz Sørensen

The Urge to Detect, the Need to Clarify

Gricean Perspectives on Information, Misinformation,
and Disinformation



Supervisor: Jens-Erik Mai

Co-supervisor: Erik J. Olsson

Submitted: 05/02/2016

Title: The Urge to Detect, the Need to Clarify

Subtitel: Gricean Perspectives on Information, Misinformation, and Disinformation

Royal School of Library and Information Science, Faculty of Humanities
University of Copenhagen

Author: Sille Obelitz Sørensen

Supervisor: Professor Jens-Erik Mai, RSLIS, University of Copenhagen

Co-supervisor: Professor Erik J. Olsson, Theoretical Philosophy, Lund University

Assessment committee:

Jack Andersen, Associate Professor PhD, RSLIS, University of Copenhagen

David Bawden, Professor of Information Science, City University London

Patrick Blackburn, Professor of Philosophy, Roskilde University

Thesis submitted Feb 5th 2016

Word count: 74.358

Contents

Contents	003
List of figures	005
Abstract	006
Dansk resumé	007
Acknowledgements	008
Introduction	010
The urge to detect	010
The need to clarify	013
Overview of dissertation	015
Ch. 1 Philosophy of Information	018
1 An informational landscape	018
1.1 Lying/Deceiving/Misleading	019
1.2 Dictionaries	023
1.3 Philosophy of Information – the project, the field	032
1.4 The method of explication	042
1.5 The Inevitable	045
1.6 Three asymmetries emerge	055

Ch. 2 Theoretical Framework	060
2 Grice – Communication as cooperation	060
2.1 Ordinary language philosophy	061
2.2 The Gricean program – meaning as intention	067
2.3 Logic and Conversation	071
2.4 Influence and critique	079
2.5 Gricean framework	081
Ch. 3 Meaning	084
3 Meaning	084
3.1 Theories of meaning	085
3.2 Objective information and the rise of meaning	088
3.3 Meaning in the Infosphere	096
3.4 Information and misinformation as communication?	098
3.5 Disinformation	102
3.6 Meaning and truth	105
Ch. 4 Truth	107
4 Truth	107
4.1 Theories of truth	108
4.2 Truth, falsity, and alethic neutrality	113
4.3 Truth and intention	133
Ch. 5 Intention	135
5 Intention	135
5.1 Theories of intention	136
5.2 Misleading – Intention and intentionality	140
5.3 Belief and interconnections	144
5.4 Intention, truth, and transitivity	147
5.5 Intention and unified conceptualization	150

Ch. 6 Unified Conceptualization	152
6 Discussions	152
6.1 The information booth	153
6.2 Semantics vs. pragmatics	156
6.3 Verbal dispute? Information and truth	160
6.4 Disinformation as a success term?	167
6.5 A unified conceptualization	168
Conclusion	179
What to detect?	179
What is clarified?	181
Perspectives	184
References	186

List of figures

Figure 2.1 – Maxims	073
Figure 6.1 – Three levels and four notions of information.	175

Abstract

The Urge to Detect, the Need to Clarify: Gricean Perspectives on Information, Misinformation, and Disinformation takes its point of departure in automatic detection of information, misinformation, and disinformation.

It argues that in order to enable automatic detection of information, misinformation, and disinformation conceptual investigation of these notions and especially their interconnections are needed. *The Urge to Detect...* provides such an investigation based on the questions:

- How have the notions of information, misinformation, and disinformation, and especially their interconnections, been conceptualized in the literature within philosophy of information?
- How do the notions of ‘meaning’, ‘truth’, and ‘intention’ shape the conceptualizations of, and affect the interconnections between, information, misinformation, and disinformation?
- How can ordinary language philosophy, esp. Paul Grice’s Cooperative Principle and theory of meaning, shape a unified conceptualization of the notions of information, misinformation, and disinformation?

The main emphasis are on accounts of information, misinformation, and disinformation, respectively, provided by philosophers and information studies scholars Dretske (1981, 1983, 2008), Floridi (2005a, 2007, 2011b), Fox (1983), and Fallis (2009, 2011, 2014, 2015). The different accounts are examined and evaluated through Carnap’s (1950) method of explication in order to develop a unified conceptualization of information, misinformation, and disinformation within a Gricean framework.

The dissertation develops a unified conceptualization of information, misinformation, and disinformation. It consists in three levels. The first level has an overall notion of “Information”- the unspecified conglomerate of all the notions of information. The second level consists in two notions of information: Natural and nonnatural information based on Grice’s distinction between natural and nonnatural meaning. The third level consists in three notions: information as intentional non-misleading, misinformation as unintended misleading, and disinformation as intentional misleading. These three notions are different aspects of nonnatural information and can be either true or false.

Dansk resumé

The Urge to Detect, the Need to Clarify: Gricean Perspectives on Information, Misinformation, and Disinformation tager sit udgangspunkt i automatisk identifikation af information, misinformation og disinformation.

Der argumenteres for, at muligheden for automatisk identifikation af information, misinformation og disinformation kræver en konceptuel undersøgelse af disse begreber med særlig vægt på deres indbyrdes relationer. *The Urge to Detect...* er en sådan undersøgelse baseret på spørgsmålene:

- Hvordan er begreberne information, misinformation og disinformation og deres indbyrdes relationer konceptualiseret i litteraturen indenfor *Philosophy of Information*?
- Hvordan former og påvirker begreberne 'mening', 'sandhed' og 'intention' konceptualiseringerne af, og de indbyrdes relationer mellem, information, misinformation og disinformation?
- Hvordan kan dagligsprogsfilosofi, særligt Paul Grice's Kooperative Princip og meningsteori, forme en samlet konceptualisering af begreberne information, misinformation og disinformation?

Hovedvægten i afhandlingen ligger på teorier om information, misinformation og disinformation, som de tager sig ud hos filosoffer og forskere i informationsstudier, særligt Dretske (1981, 1983, 2008), Floridi (2005a, 2007, 2011b), Fox (1983) og Fallis (2009, 2011, 2014, 2015). De forskellige teorier undersøges og evalueres via Carnap's (1950) '*method of explication*' – en metode hvorved begreber ekspliciteres og præciseres i henhold til dagligdags forståelser af samme – i et Griceansk perspektiv med henblik på at udvikle en samlet konceptualisering af information, misinformation og disinformation.

Afhandlingen udvikler og præsenterer en samlet konceptualisering af information, misinformation og disinformation. Denne konceptualisering består af tre niveauer. På det første niveau findes et overordnet informations-begreb "Information", der dækker over al brug af begrebet information, når dette ikke er ekspliciteret. På det andet niveau findes to informations-begreber: Naturlig information og ikke-naturlig information baseret på Grice's distinktion mellem naturlig og ikke-naturlig mening. Det tredje niveau dækker over tre begreber som aspekter af ikke-naturlig information: information som intentionel ikke-vildledning, misinformation som ikke-intenderet vildledning og disinformation som intentionel vildledning. Disse tre aspekter af ikke-naturlig information kan alle være sande eller falske.

Acknowledgements

The philosophical and scientific endeavor can be a lonely one and yet it is never conducted alone. The work presented in this thesis – as well as the work behind it – is no exception: it is the result of the collaboration of many minds.

A great many people deserves to be thanked and I would like to start with my two supervisors Prof. Jens-Erik Mai (University of Copenhagen, RSLIS) and Prof. Erik J. Olsson (Lund University, theoretical philosophy). Thank you for the discussions, the encouragement, the enthusiasm, the valuable comments, the trust in my abilities, for accepting and respecting the premises of the project, for having my back, and for genuinely caring. You have been invaluable and I could not have wished for better guidance. Thank you!

I also wish to thank Prof. Fred Adams, Prof. Don Fallis, and Prof. Gualtiero Piccinini for their kind answers to my many questions, whether in person or by email, and for their rewarding comments on whatever draft I have sent them. They have inspired and challenged my views.

Thanks to the opponents at my work-in-progress seminar Volkmar Engerer, Per Hasle, Laura Skouvig, and Peter Øhrstrøm who provided valuable suggestions and comments on the initial parts of my work. Thanks to Jesper Kallestrup, Klemens Kappel, Kay Matthiesen, and Esben Nedenskov Petersen for valuable comments and interest in my work and to Jens Ulrik Hansen, Kira Vrist Rønn, and Kristine Buch Uldahl for comments, coffee, and conversation in general.

I also wish to thank all my colleagues at RSLIS: The library, the administrative, the technical, and the scientific staff, and my PhD fellows: Ole, Sara, Rikke, Lorna, Lasse, Jesper, and Maria. Niels-Peder Osmundsen Hjøllund deserves a special thank – with you around it never actually gets lonely. I have enjoyed our many conversations about life and academia, life in academia, and the academic life.

I would also like to thank my colleagues at Dept. of Philosophy and Dept. of Arts and Cultural Sciences, Lund University: The members of the research group LUIQ (Lund University Information Quality), the members of the interdisciplinary research project *Knowledge in a Digital World* (KDW), and the participants at the Higher Seminars in Theoretical Philosophy – especially Cecilia Andersson, Jutta Haider, Sara Kjellberg, Olof Sundin, Cathrine Felix, Emmanuel Genot, George Masterton, Rasmus K. Rendsvig, Staffan Angere, Justine Jacot, Paolo Proietti, Paula Quinon, Frank Zenker, Asger Hinrup-Kirkeby, Ingvar Johansson, and Jeroen Smid.

I am grateful to the participants at the combined Dretske conference and 6th Workshop in the Philosophy of Information at Duke University, organized by Orlin Vakarelov; as well as the participants at two conferences at University of Edinburgh – *Gettier at 50* and *Extended Knowledge* – in particular Adam Carter and Duncan Pritchard for kindly offering me a seat at the philosophy department for a short research stay at the University of Edinburgh.

Thanks to staff and students in the Algorithms Reading Group at Dept. for Philosophy and Science Studies, Roskilde University.

Further I would like to thank Jack Andersen, David Bawden, and Patrick Blackburn for their willingness to participate as members of the assessment committee.

To my friends and family: Thank you for being there! Mads and Valdemar thank you for being who you are!

Sille Obelitz Søe

Copenhagen

Feb. 2016

Introduction

The urge to detect

- Information, misinformation, and disinformation

We live in the age of information. We speak of the information society, the digital information society. We have seen the rise of new digital information technology. We live with information and information technology as an indispensable part of everyday life. The Internet has become an important mean for communication, news, and information sharing in large parts of the world. Further, the information is often processed by algorithms for decision-making and as such our societies are increasingly run by algorithms. Algorithms developed on the basis of current information with the purpose of processing and controlling other information in order to 'make decisions' (Pasquale, 2015). People also use information to make decisions. However, 'information' comes in many varieties spanning from good to bad, useful to harmful, and sincere to deceptive.

"The London Eye was on fire during the 2011 England Riots! Or was it? Social networks are rife with lies and deception, half-truths and facts. But irrespective of a meme's truthfulness, the rapid spread of such information through social networks and other online media can have immediate and far-reaching consequences. In such cases, large amounts of user-generated content need to be analysed quickly, yet it is not currently possible to carry out such complex analyses in real time." (PHEME, 2014).

Information – in terms of rumors, lies, and deception – and the possible consequences when it is spread through social networks are the motivation of the PHEME-project – the lie detector to test rumors on social media. It is a three year EU funded project, which develops

technology and visual platforms to detect and categorize rumors on the social media in relation to their truth-value. The rumors are categorized into four types: speculation, controversy, misinformation, and disinformation¹ (PHEME, 2014; BBC News, Technology, 2014; Sheffield News, 2014). The project is centered on the notion of veracity (i.e. truthfulness), called the fourth challenge of Big Data (the three others being volume, velocity, and variety) (PHEME, 2014).

The motivation and challenge for the PHEME-project is almost an echo of the motivation which Karlova and Fisher (2013) present in connection with the development of their diffusion model of misinformation and disinformation, although the goal is different. Karlova and Fisher's outset is not the London riots, but instead the Fukushima radiation leak in 2011 and the economic situation in Europe in 2012:

"As gossip and rumours abound, it is difficult to distinguish among information, misinformation, and disinformation. People enjoy sharing information, especially when it is 'news'. Although they may not believe such information themselves, they take pleasure in disseminating it through their social networks. In this way, misinformation (inaccurate information) and disinformation (deceptive information) easily diffuse, over time, across social groups. Social media, such as Twitter and Facebook, have made dissemination and diffusion easier and faster." (Karlova & Fisher, 2013, p. 2).

In Karlova and Fisher's (2013) diffusion model misinformation and disinformation "extend the concept of information through their informativeness" and misinforming and disinforming function as types of information behavior.

Along the same lines Kumar and Geethakumari (2014) propose the development of an algorithm which detects misinformation and disinformation in online social networks. The example is Twitter and the detection is based on measurements of trust in, and credibility of, the original tweeter of a given tweet:

"Our methodology involves detecting cues of deception in online social networks to segregate false or misleading information with the intention of developing an effective tool for evaluating the credibility of information received by a user based on

¹ In the press-release from The University of Sheffield these categories are defined as: "speculation – such as whether interest rates might rise; controversy – as over the MMR vaccine; misinformation, where something untrue is spread unwittingly; and disinformation, where it's done with malicious intent." (Sheffield News, 2014).

the source of the message as well [as] its general acceptability in the network.”
(Kumar & Geethakumari, 2014, p. 3).

The PHEME-project, Karlova and Fisher’s (2013) diffusion model, and Kumar and Geethakumari’s (2014) algorithm-project are all examples of the (increased) interest in the potential dangers and unfortunate consequences that an open internet and social media have on the spread of misinformation and disinformation. However, in order to detect misinformation and disinformation – for example through the use of algorithms – we need to know which features to build into the algorithms; i.e. which features the algorithms ought to detect.

All three projects are based on specific notions, definitions, and understandings of misinformation and disinformation. In the PHEME-project misinformation is defined as something untrue spread unwittingly and disinformation is something untrue spread with malicious intent (Sheffield News, 2014; note 1). Karlova and Fisher (2013) define misinformation as inaccurate information and disinformation as deceptive information, whereas Kumar and Geethakumari (2014) define the two notions through Oxford Dictionary: “*Misinformation* is false or inaccurate information, especially that which is deliberately intended to deceive. *Disinformation* is false information that is intended to mislead, especially propaganda issued by a government organization to a rival power or the media.” (Kumar & Geethakumari, 2014, p. 3). Thus, although all three projects refer to the notions of misinformation and disinformation these notions are each defined in three different ways.

The differences between the definitions might seem minor, and yet, they do have various implications which turn them into completely different notions. These various implications arise due to the words used to define the notions – for instance, it makes a difference whether misinformation is defined as intended or unintended misleading and whether it is defined as false or inaccurate information. When misinformation is defined as ‘untrue’ and ‘spread unwittingly’ (Sheffield News, 2014 (PHEME)) this is the same as ‘false’ and ‘unintended’ which is quite different from misinformation as ‘false or inaccurate’ and ‘intended to deceive’ (Kumar & Geethakumari, 2014). ‘Inaccuracy’ does not entail ‘falsity’ as something true can be inaccurate if it does not provide the full picture of a given matter. Therefore, when misinformation is defined as ‘false or inaccurate’ it means that it might be false, but is not necessarily false. Thus, one definition of misinformation reads ‘necessarily false and unintended’ (i.e. PHEME), whereas the other reads ‘possibly false and intended to deceive’ (i.e. Kumar & Geethakumari, 2014). Last there is misinformation as just ‘inaccurate’ (Karlova & Fisher, 2013). As stated above, ‘inaccuracy’ does not entail ‘falsity’ but it clearly does not entail ‘truth’ either. Thus, in the definition of misinformation as simply ‘inaccurate information’ there is neither a specification of truth-value nor a specification of intentions –

two aspects which are present within the other definitions of misinformation. As regards disinformation it is also the relation to truth-values and intentions which differentiate among its definitions. Disinformation as ‘untrue’ and ‘spread with malicious intent’ (Sheffield News, 2014 (PHEME)) means that it is ‘false’ and entails some sort of intention. However, besides being ‘malicious’ it is not specified what the intent should or could be (e.g. deceiving, misleading, obstructing, etc.). Such a specification is present in the definition of disinformation as ‘false’ and ‘intended to mislead’ (Kumar & Geethakumari, 2014) although the intention here does not have to be malicious. In the last definition disinformation is just ‘deceptive information’ (Karlova & Fisher, 2013) which again means that there is no specification of truth-value.

The need to clarify

– Information, misinformation, and disinformation

The significant differences among the definitions within the three projects suggest that a better and clearer understanding of the notions is needed. What is it really that these different projects try to detect? It is clearly not the same, but should it be? What would happen if an agent employed all the different tools within the same framework – i.e. used the algorithm within Twitter and at the same time used the PHEME-platform? Different answers to the same question – i.e. is this misinformation? – would be provided. A given tweet or Facebook update that is categorized as misinformation through Kumar and Geethakumari’s algorithm would be categorized as disinformation through the PHEME-project. Add to this, that those who consult the different tools may have yet another understanding of the notions. The objective of the detecting-projects is to improve decision-making in order to avoid the bad consequences that can arise when decisions are made on the basis of bad, wrong, false, or inaccurate information (i.e. misinformation and disinformation). However, if it is unclear what is detected across the different projects (on the different platforms, through the different algorithms) then decision-making is not improved.

When it comes to decision-making it is especially the notions of misinformation and disinformation which pose a problem and it is truthful information which is strived for. Thus, the detecting-projects operate with definitions of all three notions. Furthermore, the three notions are often presented as a trichotomy – as three notions that belong together in some way or other – without clear specifications of how they are connected. Therefore, these notions must be clarified through systematic and conceptual analysis (i.e. through a philosophical framework) with special attention to their interconnections.

The more common approach is to explore one of the notions (information, misinformation, or disinformation) separately and independently of the others, which, if present at all, are only

mentioned². Often it is the notion of information which is at the center of analysis and exploration. The notions of misinformation and disinformation have not been attended to as much as the notion of information³, though systematic and conceptual work on these notions has been done (e.g. Fox, 1983; Fallis, 2009, 2014, 2015; Fetzer, 2004; Skyrms, 2010). As a consequence of the separate work on information, misinformation, and disinformation, respectively, the connections between these three notions have not been explored or attended to. Therefore, the question whether the connections between information, misinformation, and disinformation – that are suggested in the separate works on these notions – fit with the proposed definitions of the three notions, respectively, arises. In light of the scope of initiatives such as the PHEME-project, Kumar and Geethakumari’s algorithm-project, and Karlova and Fisher’s diffusion model this thesis argues that the notions of information, misinformation, and disinformation should be analyzed and discussed in connection to one another – as a whole. Furthermore, the three notions should be redefined, also in connection to one another, in order to consolidate the trichotomy that the notions are indicated to constitute – or at least to provide definitions which coherently explain the connections between the three notions in another way.

The argument is conducted on basis of the questions:

- How have the notions of information, misinformation, and disinformation, and especially their interconnections, been conceptualized in the literature within philosophy of information?
- How do the notions of ‘meaning’, ‘truth’, and ‘intention’ shape the conceptualizations of, and affect the interconnections between, information, misinformation, and disinformation?
- How can ordinary language philosophy, esp. Paul Grice’s Cooperative Principle and theory of meaning, shape a unified conceptualization of the notions of information, misinformation, and disinformation?

² With the exception of Fox (1983), who explores and analyzes both information and misinformation. However, his main emphasis is on information.

³ A quick search on Stanford Encyclopedia of Philosophy (plato.stanford.edu) reveals, not surprisingly, that the word/term information is used much more frequently than the words/terms misinformation and disinformation. Thus, by Oct. 30th 2014, a query for ‘information’ reveals 1584 documents, whereas a query for ‘misinformation’ reveals 14 documents, and a query for ‘disinformation’ only reveals one document. The sole document on ‘disinformation’ which also contains ‘misinformation’ and ‘information’ is Floridi’s *Semantic Conceptions of Information* (2005b). By Oct. 7th 2015 the figures are 1638 documents for information, 15 documents for misinformation, and 1 document for disinformation.

Overview of dissertation

The first two chapters outline the framework and the theoretical foundation for the conceptual work on the notions of information, misinformation, and disinformation. Philosophy of information is argued to be the philosophical field which ties the knot between information studies and classical philosophical fields such as epistemology and philosophy of language. Thus, the conceptual analyses are conducted within philosophy of information through Carnap's method of explication (Ch. 1) and informed by Grice's work on 'implicature', the Cooperative Principle and its maxims, as well as 'meaning' (Ch. 2). As part of the methodology for the analyses and discussions of information, misinformation, and disinformation Ch. 1 provides definitions of these notions from various English dictionaries as explicanda i.e. as a preliminary to arrive at philosophical analysis. Dretske, Floridi, Fox, and Fallis are the key authors examined and their definitions of information, misinformation, and disinformation, respectively, serve as explicata and are the definitions up for analysis and discussion throughout all chapters. Furthermore, distinctions between lying, misleading, and deceiving are explored in order to deepen the analyses of, especially, misinformation and disinformation. By the end of Ch. 1 three asymmetries in terms of 'meaning', 'truth', and 'intention' emerge. In Ch. 2 Grice's work in *Meaning* (1957) and *Logic and Conversation* (1967) is attended to in the light of its position within ordinary language philosophy. As Grice's work in some respects deviates from ordinary language philosophy aspects from the work of Austin and Frege are provided as examples of the semantics-pragmatics distinction.

In the chapters 3-5 the asymmetries are discussed and analyzed in turn. The asymmetries concern the three perspectives: 'meaning', 'truth', and 'intention'.

In Ch. 3 the notions information, misinformation, and disinformation are analyzed and discussed in connection to one another from the perspective of 'meaning'. First, it is analyzed whether 'meaning' as a concept figures in the various definitions of information, misinformation, and disinformation, and in that case which kind of meaning (i.e. a semantic notion or a pragmatic notion of meaning). Second, it is discussed how the differences in relation to 'meaning' affect the possible connections between the definitions of information, misinformation, and disinformation.

Ch. 4 is structured along the same lines, as Ch. 3, with 'truth' as the perspective for analyses and discussions of information, misinformation and disinformation. A main argument within philosophy of information is that information is truthful whereas misinformation and disinformation are characterized by falsity. This conception is challenged by Fallis' conceptual work on 'disinformation' which argues in favor of the possibility of 'true disinformation' – a notion grounded in the lying-deceiving-misleading distinctions. Thus, 'truth' as necessary condition for information and 'falsity' as necessary condition for misinformation and

disinformation are discussed. Intertwined in these discussions, with the outset of classical conceptions of 'truth', it is analyzed which notion of truth is present within the veridical definitions of information – something which rarely is explicated by the authors.

In Ch. 5 'intention', as the final perspective, which constitute an asymmetry, is at the core of analysis and discussion. The notions of intention and intentionality as discussed within philosophy of action, philosophy of intention, and the intersection between philosophy of mind and philosophy of language are presented in order to establish the differences between the two notions as well as to determine the various implications the two notions have for the definitions of which they are a part. The analyses and discussions concern the presence or absence of intentions or intentionality within the definitions of information, misinformation, and disinformation, respectively. Some definitions (i.e. those in terms of intentions and intentionality in the directedness-sense) render information, misinformation and disinformation agent-relative concepts, as intention, for instance, presupposes an active agent who can intend. Following the discussions in Ch. 4 and the collapse of the truth-false dichotomy for information vs. mis-/disinformation it is argued that it creates problems for a notion of information if intention or intentionality (in the directedness-sense) is absent. The problems, for instance, arise in cases of dissemination of true disinformation where an agent successfully misleads another agent and a third party is left with something true and misleading, where the intention to mislead is no longer present.

In Ch. 6 the strings are pulled together and discussed in order to outline how information, misinformation, and disinformation are connected to one another in accordance to all three perspectives at once (i.e. 'meaning', 'truth', and 'intention'). The various insights from the previous chapters (especially chapters 3-5) are collected and the chapter is comprised of several different discussions based on these insights. Thus, the chapter contains, but is not limited to, discussions of

- natural and nonnatural meaning
- semantics vs. pragmatics
- 'real' and 'substantial' vs. 'verbal' disputes about information
- disinformation vs. deception.

Some of these discussions are interconnected whereas others are more self-contained, albeit, they all speak to the same overall question as to 'how information, misinformation, and disinformation are connected to one another and how these connections influence their definitions'. Furthermore, how the notions information, misinformation, and disinformation are connected to one another affect how they can be and have to be defined – i.e. it affects and determines the development of a unified conceptualization of information, misinformation, and disinformation. For instance, non-misleadingness is introduced as a new perspective (a

new requirement) for the definition of information and it is argued that it is more fruitful to define information in terms of non-misleadingness instead of truthfulness. Such new perspectives have consequences for the detecting-projects and their definitions and applications of 'information', 'misinformation', and 'disinformation'.

On the basis of the various analyses and discussions in Ch.1-Ch. 6 and the insights these provide in regard to the interconnections between information, misinformation, and disinformation it is concluded that intention/intentionality and misleadingness/non-misleadingness are the distinguishing features of information, misinformation, and disinformation as the truth-false dichotomy has collapsed. Thus, truth and falsity are not sufficient features in order to automatically detect information, misinformation, and disinformation in online social network structures. Intention/intentionality and misleadingness/non-misleadingness might not easily be detected. However, Grice's Cooperative Principle and its maxims offer a heuristic which can help determine whether something is information, misinformation, or disinformation.

Ch. 1 Philosophy of Information⁴

1 An informational landscape

Questions are posed within a specific context – a specific field – and their answers must be provided within the same context or field in order to be satisfactory. Different fields enable different answers to the same questions and a field guides research (on specific questions and concepts) through the body of literature connected to that field. Thus, a wide variety of work on the notion of information has been conducted within such diverse domains as technology, physics, biology, information studies, psychology, sociology, and philosophy, etc. with completely different concepts of information as the result (Robinson & Bawden, 2014).

The scope of the detecting-projects is automatic detection of misinformation and disinformation within short semantic structures in social networks – i.e. within communication as a social practice. The detection is conducted through various information technologies such as algorithms, platforms, and apps on computers, tablets, and smartphones. The combination of communication, social networks, and information technology points to information studies as a suitable field for analyses and discussions of information, misinformation, and disinformation. However, the scope of this thesis is to analyze and discuss information, misinformation, and disinformation in connection to one another on a conceptual level in order to explore these connections and redefine the notions. The combination of communication and conceptual analysis points to analytical philosophy as a

⁴ Some of the sections in this chapter have been presented in a shorter version in Søre (submitted) *Intention and Misleading*. It concerns sections 1.2.1-1.2.2, 1.5.1-1.5.2 and 1.6.

suitable field for analyses and discussions of information, misinformation, and disinformation – more precisely it points to philosophy of language and epistemology.

Fortunately, philosophy of information is an interdisciplinary field, which, among other disciplines, encompasses both information studies and analytical philosophy (cf. section 1.3). Therefore, philosophy of information is evident/obvious as the field wherein the conceptual analyses and discussions shall be conducted. Furthermore, most of the sparse conceptual work on misinformation and disinformation already figures within the literature connected to philosophy of information – either as explicit contributions to the field (e.g. Fallis 2009, 2011, 2014, 2015; and Fetzer, 2004a, 2004b) or through extensive references from within the field (e.g. Fox 1983; and Skyrms, 2010) – and are to a great extent influenced by the literature on lying, misleading, and deceiving. The three detecting-projects are also influenced by the literature on lying, misleading, and deceiving as they, besides misinformation and disinformation, work with concepts such as rumors, speculation, trust, and credibility – all concepts which are closely connected to the notions lying, misleading, and deceiving, or can be damaged by these. Not surprisingly much conceptual work on information figures within the literature of philosophy of information as well – again either as explicit contributions to the field (e.g. Fetzer, 2004a; and Floridi, 2005a, 2005b, 2007, 2011b) or through extensive references from within the field (e.g. Dretske, 1981; and Fox, 1983).

1.1 Lying/Deceiving/Misleading

When projects like the PHEME-project, Kumar and Geethakumari's (2014) algorithm-project, and Karlova and Fisher's (2013) diffusion model are concerned with misinformation and disinformation as the concepts to detect and combat in online environments, the descriptions of misinformation and disinformation are cast in terms of 'deceiving' and 'misleading'. Thus, some clarifications of 'deceiving' and 'misleading' will be useful in order to qualify the notions and definitions of misinformation and disinformation, and maybe even information.

The notions of deceiving and misleading have received thorough attention within philosophy of language and epistemology⁵ along with the notion of lying (e.g. Adler, 1997; Fallis, 2010; Mahon, 2008; Stokke, 2013; and Webber, 2013). The thrill and the dread in conceptual philosophical work is that notions and concepts – and how they are defined – are always up for discussion. Lying, misleading, and deceiving are no exceptions. Thus, there is no consensus

⁵ There are also discussions within moral philosophy. However, they are different in scope, than what can be addressed here, as they concern moral responsibility and the moral hierarchy of wrongness (i.e. whether it is worse to lie than to mislead, for instance).

on how to define these three notions. However, debates are centered on different aspects of the definitions. In philosophy of language the discussions concern the kinds of linguistic acts which are used for lying, misleading, and deceiving, respectively, whereas epistemology is mostly concerned with questions about the damage of credibility caused by lying, misleading, and deceiving in, for example, testimony. Of vital importance to the debates within the literature on lying, misleading, and deceiving are some of H.P. Grice's ordinary language philosophic work – that is, his Cooperative Principle, its maxims, and the notion of implicature (especially conversational implicature) (1967).

Understandings and definitions of lying, misleading, and deceiving are dependent on whether the theory works with a 'lying-misleading' distinction (Stokke, 2013; and Webber, 2013) or a 'lying-deceiving' distinction (Adler, 1997). In both distinctions 'lying' is defined as an act of assertion, where the liar says something which he believes to be false. The assertion of something believed-false is then distinguished from the use of false conversational implicatures⁶ where what is literally said is true, but what is implicated is false. Thus, in 'misleading' or 'deceiving' (dependent on which distinction is at play) nothing false is actually said and the utterance which carries the false implicature is not believed to be false either. In the lying-misleading distinction it is 'misleading' which is defined as false conversational implicatures and both lying and misleading are kinds of deception (Stokke, 2013; and Webber, 2013). In the lying-deceiving distinction it is 'deceiving' which is cast in terms of conversational implicature (Adler, 1997).

1.1.1 Lying

Although the notions of misinformation and disinformation are not cast in terms of lying, 'lying' is the notion against which 'misleading' and 'deceiving' are defined. Thus, some insights to the notions of misinformation and disinformation might be gained from a clarification of the notion of lying as different from 'misleading' and 'deceiving'. The standard, or most common, definition of lying is that a person lies if he says (asserts) something which he believes is false and he has the intention to deceive whomever he talks to (Adler, 1997; Mahon, 2008; Fallis, 2010; Stokke, 2013; and Webber, 2013).

⁶ An implicature (Grice, 1967) is that content of an utterance which is not literally said but which it is possible to *work out* from the way the utterance is uttered (cf. sections 2.3.2-2.3.3).

The standard definition of lying can be stated more formally as:

“To lie =_{df} to make a believed-false statement to another person with the intention that that other person believe that statement to be true.” (Mahon, 2008, section 1)

with the further specification that “[t]he condition that the statement be *believed* to be false is not the condition that the statement *be* false.” (Mahon, 2008, section 1.2). This means that it is possible to lie to someone by saying something which happens to be true as long as the liar believes it to be false and intends to deceive the addressee(s). The requirement for an intention-to-deceive, in the standard definition, means that “lying necessarily involves an intention to communicate with another person by means of a statement. It is not possible to lie to those (...) who (or which) cannot understand statements made to them.” (Mahon, 2008, section 1.4). However, the intention-to-deceive condition for lying has been challenged by the emergence of ‘bald-faced lies’ as accepted instances of lying. ‘Bald-faced lies’ are lies that everyone (liar and addressees) knows to be lies wherefore they are not intended to deceive or mislead anybody. The classic example of ‘bald-faced lies’ are witnesses who lie on stand in the courtroom in order to escape repercussions from the defendant, even though the evidence is conclusive (Fallis, 2010; Mahon, 2008; and Stokke, 2013).

Even though the intention-to-deceive is not present in instances of ‘bald-faced lies’ the intention to communicate is preserved. Different attempts have been made in order to improve the definition of lying. When all the variations are taken aside, the key features of lying are that to lie is to say something which one believes to be false (cf. Mahon, 2008). Even though it is not a requirement that lies are intended to deceive (cf. ‘bald-faced lies’), a prototypical instance of a lie is a believed-false assertion intended to deceive.

1.1.2 Misleading and deceiving

To mislead and to deceive are not necessarily linguistic acts the same way as lying is as misleading and deceiving are not bound to assertion but to actions of any kind. As long as it is an action (even an act of omission) it can be misleading or deceiving. Thus, it is possible to mislead or deceive through other means than language; for example through manipulated pictures, by wearing a police uniform, when one is not a policeman, by gestures, or the like. However, it is possible to mislead and deceive verbally (i.e. through linguistic acts) without lying. Verbal misleading and deceiving – when not pure instances of lying – are exercised through conversational implicatures, more precisely through conversational implicatures where what is literally said is true, but what is implicated is false (referred to as false

implicatures). An often used example⁷ of a false implicature is a guy who is approached by a murderer who wants to kill Joe. The murderer asks the guy if he knows where Joe is. The guy happens to know that Joe is in the basement but he does not want Joe to get killed. The guy also knows that the Nevada is Joe's local and often visited diner, therefore he truthfully says:

"Joe has been hanging around the Nevada a lot."

in order to get the murderer to falsely believe that Joe is at the Nevada now. Instead of just lying to the murderer and say "I don't know" it takes an effort to generate a false implicature – the utterance has to be literally true, while, at the same time, be likely to cause the addressee (here the murderer) to obtain a false belief about a certain matter (Adler, 1997; Fallis, 2010; Stokke, 2013; and Webber, 2013). Therefore, misleading and deceiving through false conversational implicatures are always intentional in the literature on lying, misleading, and deceiving.

The differences between misleading, deceiving, and lying depend on whether one adheres to the lying-misleading distinction or the lying-deceiving distinction. In the lying-misleading distinction both lying and misleading are kinds of deception (with exception of 'bald-faced lies' which are not deceptive) (Stokke, 2013; and Webber, 2013). In the lying-deceiving distinction lying entails the intention to mislead and deceive (Adler, 1997). In both distinctions misleading and deceiving are intentional and even though lying does not necessarily entail the intention to deceive (cf. 'bald-faced lies'), the provision of a believed-false assertion is intentional and not a mere mistake. Thus, the different definitions of lying, misleading, and deceiving do not account for un-intended misleading. One exception is the document *The Definition of Lying and Deception* on Stanford Encyclopedia of Philosophy (2008) where (all the variations taken aside) to lie is to say something which one believes to be false, deception is intended to cause the addressee to hold false beliefs by linguistic or non-linguistic means, and misleading is when someone unintentionally causes another person to hold a false belief (Mahon, 2008).

The key features, which are important to keep in mind, are that: 'lying' is the assertion of something believed-false, which can be either deceptive or non-deceptive (e.g. 'bald-faced lies'), and they are told to an addressee with the intention that he will gain a false belief; deception is always intentional and can be brought about by different kinds of actions both linguistic (i.e. through false implicatures and lies) and non-linguistic, and even by acts of omission, as when someone withholds information and thereby causes someone else to gain a

⁷ The example is based on the example from Adler (1997). A similar Nevada-example is also used in Fallis (2014).

false belief; misleading is most often defined as caused by the use of false conversational implicatures and is also described as intentional in the lying/misleading/deceiving literature. However, in Mahon (2008) there is the possibility that mere ‘misleading’ – i.e. when it is not an ingredient in ‘lying’ and ‘deceiving’ – might be un-intentional. Furthermore, ‘lying’ is not a success term. This means that the act of lying is not dependent on whether or not someone is actually misled or deceived. Deception on the other hand is a success term, wherefore deception can only take place if someone is actually deceived. If no one is deceived it is a failed attempt of deception.

The distinctions between ‘lying’, ‘misleading’, and ‘deceiving’ are important as the notions of misinformation and disinformation as used in the PHEME-project (2014), Karlova and Fisher’s diffusion model (2013), and Kumar and Geethakumari’s algorithm-project (2014) are cast in terms of misleading and deceptive information along with ‘inaccuracy’ and ‘falseness’. Thus, the lying/misleading/deceiving-distinction is one step in the qualification of the notions misinformation, disinformation, and information. Another, equally important, step is to get a better and deeper sense of the definitions or basic understandings of the notions information, misinformation, and disinformation as this might help to grasp why and how the ‘detecting-projects’ can employ such different notions of information, misinformation, and disinformation (cf. Introduction).

1.2 Dictionaries

When in need of a definition or a basic understanding of some notion or concept the quickest and most common place to look first is in dictionaries and encyclopedias.

Depending on the dictionary and its scope the definitions give an indication of the various usages or normative definitions⁸ of a given notion in common language – from the historical development, to word origin and etymology, to specific usages within different disciplines. Where a “synchronic dictionary presents the reader with a snapshot of a large number of related lexical strands at one point in time (typically the present)” (Simpson, 2013, p. 342) a diachronic dictionary “address[...] language across its history.” (Simpson, 2013, p. 341). Thus, “[t]he majority of the text in a historical dictionary such as the *OED* [Oxford English Dictionary] consists not of definitions or etymologies, but of historical and modern illustrative

⁸ The online dictionaries do not always make it clear which view of language they employ, however, it seems that both Oxford Dictionaries and Cambridge English employ a use-centered view of language (<http://www.oxforddictionaries.com/words/about>; <http://www.cambridge.org/dk/cambridgeenglish/about-cambridge-english/cambridge-english-corpus>).

quotations documenting the use of senses of words over time from their earliest recorded use.” (Simpson, 2013, p. 345). Dictionaries and encyclopedias are classically seen as repositories of knowledge – knowledge ripe to pick and use for whatever need people might have when it comes to words, notions, and concepts (cf. Simpson, 2013; and Merriam-Webster, 1991, preface). “Readers make the reasonable assumption that dictionaries are authoritative. (...) But almost all editorial text in a dictionary is relative, not absolute.” (Simpson, 2013, p. 345) it is dependent on the lexicographers, the language-view, and the scope of the dictionary, etc.

In detecting-projects, such as the PHEME-project and the algorithm-project (Kumar & Geethakumari, 2014; cf. Introduction), dictionaries often serve as this repository of ready-made definitions. The notions of misinformation and disinformation, as well as information, are used in commonsense-interpretations which resemble the senses found in dictionaries slightly changed to fit the overall purpose of the project (e.g. PHEME) or they are adopted straight out of the dictionary (e.g. Kumar and Geethakumari, 2014). Hence, PHEME defines misinformation within a Twitter-context as “something untrue (...) spread unwittingly” (BBC News, 2014) and disinformation within the same context as “false information (...) spread with malicious intent” (BBC News, 2014; cf. Introduction). Both definitions resemble the definitions in various English dictionaries (cf. section 1.2.1) except they are adapted to the social media context Twitter by use of the word ‘spread’. Kumar and Geethakumari (2014) simply cite the Oxford English Dictionary for their definitions. In connection to the notion of misinformation Karlova and Fisher (2013) write:

“[u]nfortunately, misinformation does not seem to earn the attention it deserves. While the Oxford English Dictionary defines misinformation as, ‘*wrong or misleading information*’, few authors have discussed the topic in detail. Authors commonly cite the OED definition without further analysis or discussion (...)” (Karlova & Fisher, 2013, p. 2).

Furthermore, philosophical studies of specific notions and concepts sometimes use dictionaries or common-sense descriptions as well in order to capture the notion in its ‘ordinary sense’. For instance, C.F. Fox and philosopher Fred Dretske both take their point of departure in “the ordinary language notion of information” (Fox, 1983, p. 5), “the nuclear sense of the term “information”” (Dretske, 1981, p. 45), and “information as it is ordinarily understood” (Dretske, 1983, p. 56) in their philosophical studies of the notion of information. However, Fox (1983) and Dretske (1981, 1983) end up with completely different notions of information (cf. sections 1.5.1 and 1.5.2).

Thus, in order to understand how the ‘ordinary sense’ of information can result in completely different philosophical accounts as well as why the definitions within the detecting-projects

differ in various aspects, definitions of the notions information, misinformation, and disinformation as they are found in a selection of English dictionaries will be presented. This presentation serves two goals: first, it will give a greater sense of how the notions are used within detecting-projects, i.e. what the algorithms are supposed to detect; second, the dictionary-senses of information, misinformation, and disinformation will serve as a preliminary basis (i.e. as explicanda, cf. section 1.4) for the discussions of the philosophical accounts of these three notions.

1.2.1 Misinformation and disinformation

The scope of the detecting-projects is to detect disinformation and/or misinformation in social networks as these notions are associated with falsity, misleadingness, wrongness, deception, and the like, and are thought to have a negative effect on, for example, decision-making. Although the PHEME-project (2014), the algorithm-project (Kumar & Geethakumari, 2014), and the diffusion model (Karlova & Fisher, 2013) employ different understandings and definitions of misinformation and disinformation, respectively, and although these differences may have unfortunate consequences for agents who consult more than one platform (cf. Introduction), all the definitions bear resemblance to definitions of misinformation and disinformation from various English dictionaries.

In Oxford English Dictionary online misinformation is defined as “[w]rong or misleading information”, whereas it is defined as “wrong, bad, opposite to or lack of information” in Merriam-Webster (through Britannica.com). The definitions of misinformation from the PHEME-project (i.e. something untrue) and the diffusion model (i.e. inaccurate information) bear some resemblance to the definitions from OED and Merriam-Webster, although neither OED nor Merriam-Webster define misinformation in terms of falsity, untruth, or inaccuracy. However, when the definitions of ‘misinform’ from Dictionary.com are taken into account the resemblance is apparent. To misinform is “to give false or misleading information to” or “to give incorrect information to” (Dictionary.com) someone or somebody. From these definitions three general observations can be made: first, misinformation is defined in terms of information, as if misinformation is a kind of information; second, misinformation can be false, wrong, bad, or misleading which indicate that misinformation does not have to be false as long as it is misleading, or inaccurate, or something else along the same lines; third, no intentions on the behalf of the misleader are mentioned. This might point to a conception of ‘misleading’ as unintended in the same sense as found in Mahon (2008) (cf. section 1.1.2).

When it comes to disinformation the resemblance, between the various definitions from the detecting-projects (cf. Introduction) and the dictionaries, is even closer than the resemblance between the definitions of misinformation. Disinformation is defined in terms of deliberate falsity, deliberate spread, intention, and deception – terms that reappear in the definitions

from the detecting-projects. More specifically, disinformation is defined as “false information intended to deceive or mislead” (Dictionary.com), “false information deliberately and often covertly spread (as by the planting of rumors) in order to influence public opinion or obscure the truth.” (Merriam-Webster through Britannica.com), and “[t]he dissemination of deliberately false information, esp. when supplied by a government or its agent to a foreign power or to the media, with the intention of influencing the policies or opinions of those who receive it; false information so supplied.” (Oxford English Dictionary online).

There are also some general observations to be made from the definitions of disinformation. First, in all the definitions it is emphasized that disinformation is false information, which indicates that disinformation is a kind of information. Second, disinformation is defined as provided deliberately with the intention to gain influence or with the intention to deceive. The combination of the falseness and the deliberate dissemination or intention to deceive, points to why it is appealing to be able to detect disinformation on the web. A third observation also concerns the required falseness of disinformation. In the distinctions between lying, misleading, and deceiving one general agreement is that only lying requires the assertion of something believed-false (NB actual falseness is not a requirement for lying as long as the liar believes the content of the utterance to be false). Deceiving and misleading – which are the notions used to define disinformation – do not have a requirement for believed falseness and they are not bound to utterances. In ‘misleading’ and ‘deceiving’ it is the implicature and thereby that which is implicated, which has to be false (cf. sections 1.1.1 and 1.1.2). As it is not specified in the definitions from the dictionaries which form the false information must or can have in order to constitute disinformation (e.g. is it only utterances which count? Or does the ‘disinformation’ also apply to written text, gestures, doctored photographs, forged maps, etc.?) it is not possible to establish whether ‘to disinform’ is a kind of lying or a kind of deceiving in general.

The distinctions between deceiving and misleading, especially whether or not something is intended to mislead, seem to be the main differences between misinformation and disinformation in the definitions examined so far. However, there is a disagreement between the dictionaries as to how to define misinformation. The Oxford English Dictionary online, Merriam-Webster, and Dictionary.com define misinformation and disinformation as two distinct notions, whereas, for example Oxford Dictionaries online, defines the notions as synonyms with misinformation defined as “[f]alse or inaccurate information, especially that which is deliberately intended to deceive” (i.e. the definition of misinformation that is used in the algorithm-project). Definitions of misinformation along the same lines – that is, false information which is intended to deceive – are found at collinsdictionary.com, dictionary.cambridge.org, and macmillandictionary.com. These disagreements as to whether or not misinformation is intended to deceive mirror the disagreements in the literature on

lying, misleading, and deceiving as to whether 'misleading' is intentional or can be unintentional (cf. section 1.1.2).

At Dictionary.com, one of the dictionaries which do distinguish between misinformation and disinformation, it is actually noted in relation to the definitions of both notions that they can be confused⁹. This possibility of 'confusion' between the meanings of the words misinformation and disinformation may derive from the relation between disinformation and the Russian word Dezinformatsiya. According to *The Merriam-Webster New Book of Word Histories* (1991) Dezinformatsiya "is purported to have been the name of a department of the KGB formed in 1955, which oversaw the dispensing of propaganda to international media and government organizations." (Merriam-Webster, 1991, p. 144). In Russian 'dezinformatsiya' means misinformation (Merriam-Webster, 1991), so it seems that the Russians have taken the word they had, i.e. dezinformatsiya (misinformation), and used it for a specific purpose (now called disinformation) whereby the transformation into the English word disinformation adopted this specific use as its definition, the deliberate and intended dissemination of inaccurate or false 'information' in order to mislead or deceive, especially foreign governments or the media. However, *The Merriam-Webster New Book of Word Histories* (1991) points out that it is a matter open to speculation whether the word disinformation is actually derived from dezinformatsiya as there are records of a German 'Disinformation Service' as early as 1939 (Merriam-Webster, 1991).

The variations between the definitions of disinformation are minor and mostly concern the actual words used in the descriptions. Thus, the essence of disinformation seems to be deliberate dissemination of false information. In some sense the same goes for misinformation albeit there are two different tracks. One track is those dictionaries which define misinformation as false, wrong, or misleading information – with the addition of 'inaccuracy' when 'to misinform' is taken into account. The variations between these definitions are minor and only concern the words used for description. The other track is those dictionaries which define misinformation as a synonym for disinformation. In this track misinformation is defined in the same terms as disinformation i.e. as intended deception and the variations between these definitions are also minor.

At vocabulary.com another kind of definition is present. Misinformation and disinformation are seen in relation to one another as well as in relation to information. Disinformation is defined as "misinformation that is deliberately disseminated in order to influence or confuse

⁹ This might indicate a warning as not to confuse the notions.

rivals (foreign enemies or business competitors etc.)” and it is stated as a type of misinformation. Misinformation is defined as “information that is incorrect” and it is stated as a type of information. Information is defined as “a message received and understood”¹⁰ (Vocabulary.com). Thus, at Vocabulary.com the three notions information, misinformation, and disinformation are explicitly connected to one another – they are all information, misinformation is then a specific type of information, and disinformation is a specific type of misinformation. Besides these explicit connections, the definitions of misinformation and disinformation are similar to the definitions from the other dictionaries and share the same essences in terms of inaccurate information (misinformation) and deliberate dissemination of inaccurate information.

All the differences taken aside, there is one aspect which all the definitions of both misinformation and disinformation share without exception – namely, the reference to information.

1.2.2 Information

In contrast to the notions of misinformation and disinformation, the notion of information has multiple senses divided into categories, subcategories, and types across the dictionaries¹¹. The way in which the different senses are divided and defined as well as the number of categories, subcategories, and types varies between the different dictionaries. The variations are minor and mostly concern the categorization; there is little variation among the senses themselves, which resemble each other throughout the dictionaries¹².

In Oxford English Dictionary online the multiple senses of information are presented in three categories (I, II, III) divided into seven subcategories (1-7), each with various types (a-e). For the purposes here only the first category I. “The imparting of knowledge in general” is of interest. A selection of subcategories and types are, 1.a: “The shaping of the mind or character; communication of instructive knowledge; education, training;”, 2.a: “Knowledge communicated concerning some particular fact, subject, or event; that of which one is apprised or told; intelligence, news.”, 2.c: “As a mathematically defined quantity divorced

¹⁰ This definition of information on Vocabulary.com is one category out of three (cf. section 1.2.2).

¹¹ Due to the huge amount of different senses of information this section will not be an exhaustive presentation of all categories, subcategories, and types. Only a selection of the senses found relevant for the scope of this dissertation will be presented. For example the definition of information in law i.e. the formal accusation of a crime, which is present throughout all the dictionaries mentioned will not be taken into account. Furthermore, senses of information no longer in use or only rarely in use will not be considered either.

¹² The most common variation is between categories and subcategories i.e. some definitions or senses of information are presented as category in one dictionary but as subcategory in another dictionary.

from any concept of news or meaning;" which includes the notion of information in Information Theory, and 2.e: "contrasted with *data*: that which is obtained by the processing of data." (Oxford English Dictionary online).

In Oxford Dictionaries Online there are also multiple senses of information where category 1 is "Facts provided or learned about something or someone" and category 2 is "What is conveyed or represented by a particular arrangement or sequence of things: *genetically transmitted information*" which includes type 2.1 concerning *Computing* "Data as processed, stored or transmitted by a computer." (Oxford Dictionaries Online). The various definitions on Dictionary.com are along the same lines. In all these senses information is connected with knowledge in some way or other or is presented as a commodity which can be processed, stored, or transmitted as is seen in the mathematical or computing senses.

These senses are also present in the Learner's Dictionary by Merriam-Webster as is seen in the definition "Knowledge that you get about someone or something: facts or details about a subject", but in the examples something else is present as well: "He gave the police false/misleading *information* about his background", "We can't make a decision until we have more/further/additional *information*.", "He's accused of withholding useful/valuable/vital *information*.", "My information is [= I have been told, I understand] that he will be arriving early this evening." (Learner's Dictionary, Merriam-Webster). In these examples, besides being connected to knowledge, information can be false and misleading as well as subjective, in the sense 'My information'. As falsity and misleadingness cannot be connected to knowledge (because knowledge requires truth) these examples show an ambiguity in the notion of information; when defined by itself information is connected to knowledge, but in the definitions of misinformation and disinformation information is referred to in a sense where it can be false and misleading (cf. section 1.2.1).

At Vocabulary.com the multiple definitions of information are divided into three categories, which resemble three of the subcategories on Oxford English Dictionary online. Two of these are 1: "Knowledge acquired through study or experience or instruction" and 2: "(communication theory) a numerical measure of the uncertainty of an outcome" (Vocabulary.com). Category 1 is divided into three subcategories, whereas category 2 and 3 do not have subcategories. The three subcategories are 1 itself as well as "a collection of facts from which conclusions may be drawn" and "a message received and understood". Each subcategory is then divided into various types. As mentioned, in section 1.2.1, on

Vocabulary.com misinformation¹³ is defined as a type of information in the sense “a message received and understood” and disinformation¹⁴ is defined as a type of misinformation. Disinformation, however, is also defined as a type of information in the same sense as misinformation along with fact¹⁵ among several others. These different types as well as the broad subcategory “a message received and understood” itself indicate, in connection with the examples from the Learner’s Dictionary, a notion of information, where information is not necessarily true.

The notion of information is an ambiguous notion. When defined on its own it can be anything from a mathematical quantity devoid of agents to something which agents use in communication and learning to obtain knowledge – e.g. intelligence, fact, news, a message received and understood, etc. Furthermore, information is used in connection with falsity and misleadingness in order to define misinformation and disinformation. It seems that information can be almost anything which agents want it to be, whereas misinformation and disinformation are more fixed in their definitions as false, wrong, or bad information (misinformation) and false information intended to deceive (disinformation). The ambiguity of the notion of information provides an answer to the question how the ‘ordinary sense’ of information as a starting point can lead to completely different philosophical accounts of the notion (cf. Dretske, 1981; and Fox, 1983; cf. section 1.2) – information does not have one but multiple ‘ordinary senses’ which point in different directions. Thus, it is possible to pick an ‘ordinary sense’ of information which fits the overall goal and purpose of the desired philosophical account – for instance, the purpose can be to develop a philosophical account of information in connection to communication, learning, transmissibility, cognition, or something else entirely.

The ambiguity with regard to information is absent when it comes to misinformation and disinformation as their senses are more fixed. This suggests that misinformation and disinformation are less complex notions than information. However, the reduction of misinformation and disinformation to information which is false, misleading, and/or deceptive calls for an explication of the notion of information used in these definitions – i.e. which of the ‘ordinary senses’ of information is present within the definitions of misinformation and disinformation – wherefore the ambiguity is implicitly present within the

¹³ “information that is incorrect” (Vocabulary.com).

¹⁴ “misinformation that is deliberately disseminated in order to influence or confuse rivals (foreign enemies or business competitors etc.)” (Vocabulary.com).

¹⁵ “a statement or assertion of verified information about something that is the case or has happened” (Vocabulary.com).

definitions of misinformation and disinformation¹⁶ as well. Is it for example the notion of information in a learning perspective, a communication perspective, a cognition perspective, or as a mathematical quantity which is present within the definitions of misinformation and disinformation?

As mentioned above this presentation of misinformation, disinformation, and information as defined in various English dictionaries gives a sense of what the detecting-projects are trying to detect. Or rather, it gives a sense of why this might prove a more complicated task than first imagined. The ambiguity of 'information' and the influence it has on the connections between information, misinformation, and disinformation pose a lot of questions as to which features to detect – For example, does misinformation have to be treated as all that which is false? If the distinction between misinformation and disinformation is deception and intention how does one detect that? Is disinformation a kind of misinformation? When it comes to 'information' what should be detected for – facts?, messages received and understood?, communication in general?, mathematical quantities?, etc. Furthermore, when different understandings are employed by the different detecting-projects problems may arise if the tools and platforms are used simultaneously (cf. Introduction).

Furthermore, the notion 'misinformation' and the verb 'misinform' can only be applied retrospectively. An agent who misinforms never would nor could say 'I have some misinformation for you'. Misinformation is unintended, wherefore the agent cannot know beforehand that it is misinformation – if an agent passes on a piece of misinformation which he knows to be misinformation then it is not misinformation anymore as the 'un-intended' aspect evaporates and it becomes disinformation instead. It is possible to say 'I was misinformed' or 'It is/was/turned out to be misinformation' but only because it is now known (i.e. retrospectively) that the proposition in question is false or inaccurate and misleading. Dependent on an agent's own knowledge about some state of affair (e.g. that p) it is also possible for that agent to describe some other agent as misinformed (e.g. if he believes $\neg p$) or as having received misinformation. When an agent misinforms or provides misinformation by stating p the receiver might be in a position to know that it is misinformation, because he has true information that $\neg p$ or because he works out the false implicature to be a false implicature. However, the notion 'misinformation' can first be applied after p is stated. The notion 'disinformation' can also first be applied retrospectively by the receiver whereas the agent, which provides the disinformation (the agent who disinforms), knows that it is

¹⁶ With the exception of Vocabulary.com where it is specified that information in connection to misinformation and disinformation is to be understood as "a message received and understood." (Vocabulary.com).

disinformation in advance. 'Information' is the single notion which can be used beforehand by the informer and the receiver insofar information is not defined by truth. If information is defined by truth, then the notion can only be applied if the information is also known – i.e. known to be true. Thus, it seems that on a veridical account of information agents can only speak of information when they have knowledge.

Therefore, the notions of information, misinformation, and disinformation and the connections between them call for a better understanding and need scrutiny. Dictionaries cannot provide an answer to how the three notions are connected. The question, conceptual in nature, demands philosophical consideration as it ties into the underlying philosophical questions 'What is information?', 'What is misinformation?', and 'What is disinformation?'.

1.3 Philosophy of Information – the project, the field

Philosophy of information is the interdisciplinary field which has the study of information as its main objective. As information is an ambiguous notion with multiple different definitions used for different purposes philosophy of information entertain a wide variation of approaches among its researchers. Within philosophy of information the ties to fields such as computer science, cognitive science, artificial intelligence, philosophy, information studies, and the like are strong and the legacy from these fields is apparent. However, the field is in the eyes of the beholder and any philosophical study, analysis, discussion, and/or investigation with information as the main concept or object is philosophy of information or at least a special variety of philosophy of information. This point is emphasized by the members of the Society for the Philosophy of Information:

"The philosophy of information is the branch of philosophy devoted to the thematic study of information in all its forms, and to the application of informational methods to new and traditional philosophical problems. The philosophy of information is not limited to any particular doctrine or methodology; rather, it is unified by its central focus on information as it plays out in both theory and practice."

(<http://www.socphilinfo.org/about-pi/about-pi>).

Thus, philosophy of information asks many of the same questions as other classical philosophical fields. These questions regard knowledge, meaning, and ethics, to name a few. However, what makes philosophy of information an independent and autonomous field and not just a kind of epistemology, philosophy of language, or moral philosophy, is that it has "its own *topics* (problems, phenomena), *methods* (techniques, approaches) and *theories* (hypothesis, accounts)." (Adams & de Moraes, forthcoming) and "is able to appropriate an explicit, clear, and precise interpretation (...) of the classic "ti esti," thus presenting itself as a specific "philosophy of"" (Floridi, 2002b, p. 135). Furthermore, it has proved strong and rich

enough to attract researchers, encompass subfields, and “withstand centrifugal forces” (Floridi, 2002b, p. 135).

To freely rephrase the first lines of the Society for the Philosophy of Information quote above with the words from Adams and De Moraes: “Philosophy of Information [is] an interdisciplinary and autonomous area of research in Philosophy”, which is also a branch of philosophy (Adams & de Moraes, forthcoming). As an alternative to philosophy of information as an interdisciplinary branch of philosophy, philosophy of information can be seen as an interdisciplinary field comprised of intersections between the various disciplines. Thus, a specific area of philosophy of information is the intersection between information studies and analytical philosophy – especially epistemology and philosophy of language. Analytical philosophy is characterized by a focus on logic, language, and conceptual analysis¹⁷ as the means to solve philosophical problems and it is these means which are most commonly used by scholars and philosophers who work in the intersection between information studies and philosophy (cf. Fallis, 2009, 2014; Floridi, 2002a, 2002b; Furner, 2004; Fox, 1983; and Mai, 2013). Hence, philosophy of information, in the intersection between analytical philosophy and information studies, engages in questions ordinarily found in information studies with regard to what information is, how it is or can be used, its relation to knowledge, communication, and learning, questions about privacy and information ethics, and what it means to live in the information age with information technology, internet, and social media. The questions asked about the nature of information etc. are asked in a way similar to the way epistemology asks questions about the nature of knowledge and its possibility and the way philosophy of language asks questions about the nature of meaning and language.

Philosophy of information, specifically the intersection between information studies and analytical philosophy, offers a framework for conceptual work on information and related concepts with a close eye to practice and application. In this interpretation of the field Jaako Hintikka’s article *Epistemology Without Knowledge and Without Belief* (2007) can be seen as an instance of philosophy of information. Although, Hintikka himself labels it Socratic Epistemology, the core in his work is the introduction of ‘information’ as the primary, or more fundamental, concept worth of investigation within epistemology. Thus, Hintikka shifts the focus from knowledge to information in the development of a new method of knowledge-seeking by questioning (Hintikka, 2007). Dretske (1981) also moves from knowledge to

¹⁷ “Conceptual analysis emerged as the primary method used by philosophers in the particular paradigm (“analytic” philosophy) that rose to dominance in Anglophone countries in the twentieth century.” (Furner, 2004, p. 429).

information, although with a different purpose and outcome than Hintikka: “Dretske was calling for a reorientation in epistemology. He did not think that epistemologists should spend their time on little puzzles or on rehashing ancient arguments about skepticism. Rather, he held that epistemology would be better served by studying the flow of information.” (Skrirms, 2010, p. 33). In the interpretation of philosophy of information as an interdisciplinary field comprised of various intersections, each intersection can be approached from both sides of the intersection – that is, the present intersection can be approached both from analytical philosophy and from information studies. Some examples of information studies scholars who work within the analytical philosophy/information studies intersection are Furner (2004) and Mai (2013) – who are both influenced by philosophy of language – as well as Fallis (2009) and Fox (1983) (other examples are Blair, 1992, 2003; Capurro & Hjørland, 2003; Cornelius, 1996, 2002; Day, 2001; Wilson, 1977). Furner (2004) argues why and how the notion of information can be disregarded in information studies in favor of more precise and useful notions from philosophy of language, whereas, Mai (2013) discusses and develops the notion of information through a pragmatic philosophy of language in order to enter the debate on information quality. Information quality, and especially the challenges opposed to it by new information technologies, is the outset for Fallis’ conceptual analysis of disinformation (2009). Fallis’ argumentation is that a better understanding of ‘disinformation’ is needed in order to detect its presence online. The method of conceptual analysis is shared by Fox (1983) in his work on the notions of information and misinformation within communication.

Furthermore, in the article *Philosophy and Information Studies* (2010), Furner investigates how philosophy and information studies inform and influence each other. In other words, he investigates the intersection thus highlighting where the philosophy of information plays out. Furner’s (2010) conclusion is that it is mostly information studies which are influenced by philosophy, especially by epistemology and philosophy of language, whereas work from information studies is hardly present within philosophy. One exception seems to be Luciano Floridi, one of the main philosophers within philosophy of information.

Floridi (2002a) acknowledges the relation between philosophy and information studies – or library and information science (LIS) – and writes that “[t]he relation between the two fields [LIS and philosophy] seems intuitive and undeniable, but specifying its precise nature has proved a complex and controversial task.” (2002a, p. 38). In the interpretation above this relation is defined as the intersection between analytical philosophy and information studies as one of many intersections with the label philosophy of information. However, Floridi describes another relation between the two fields which stems from his own conception of philosophy of information. He argues that, instead of social epistemology, philosophy of information should serve as the theoretical foundation for library and information science

which then could be seen as applied philosophy of information; LIS would be considered to be a branch of philosophy of information (Floridi, 2002a).

Floridi's conception of philosophy of information is more of a world view than a bunch of intersections between different fields or disciplines which establishes a field of its own (e.g. analytical philosophy and information studies as a part of the interdisciplinary field philosophy of information). However, Floridi is one of the main philosophers within philosophy of information regardless of the differences among interpretations. Whether philosophy of information is considered

- an interdisciplinary branch of philosophy;
- an interdisciplinary field comprised of intersections between disciplines; or
- a foundational world view

Floridi is often the outset when philosophers and researchers explicitly address philosophy of information (cf. Fallis, 2011; Fetzer, 2004a; Furner, 2010; Mai, 2013; Scarantino & Piccinini, 2010; Van der Veer Martens, 2015). In other words, when it comes to philosophy of information Floridi's work has become the point of departure whether one agrees or disagrees with his views. Therefore, some insight to philosophy of information as conceived by Floridi is necessary in order to understand what is at play in the various theories which address 'philosophy of information' in connection to Floridi.

1.3.1 PI in Floridi's interpretation

Luciano Floridi describes philosophy of information as a foundational philosophy due to its main objective – the question "what is information?". According to Floridi, information is the most foundational concept, more foundational than knowledge, belief, mind, or life itself (Floridi, 2002b). When information is the most foundational concept it follows that information cannot be reduced and explained in terms of other concepts. Instead, the idea is that other concepts such as knowledge, meaning, and mind, etc. can be explained in informational terms. To reduce certain concepts to information in order to explain them is not a new approach. In fact this was exactly what Fred Dretske did in *Knowledge and the Flow of Information* (1981) where he provided an informational account of knowledge. However, what is fairly new is the formulation of philosophy of information as a field in its own rights. In Floridi's words "[t]he philosophy of information (PI) is the philosophical field concerned

with (a) the critical investigation¹⁸ of the conceptual nature and basic principles of information, including its dynamics, utilization and sciences, and (b) the elaboration and application of information-theoretic and computational methodologies to philosophical problems.”(Floridi in Π research network, 2013, p. 32, slightly rephrased from Floridi, 2002b, p. 137). Thus, philosophy of information has two scopes: (a) information as the most foundational concept – that is an ‘information first’-approach¹⁹; (b) innovation in terms of the development of an information-theoretic philosophical method.

Within philosophy of information, as formulated by Floridi, there are four important concepts:

- The information revolution
- The infosphere
- Inforgs
- The “onlife”

The information revolution is the era we live in. It is determined by the internet and the dominance of ICT's in everyday life. It has changed our way of living and interacting with the world. It is a new circumstance from which it is impossible to go back. The infosphere is the semantic environment which humans inhabit due to the information revolution. Through the internet and internet-based services as e-mail, Facebook, Google, etc. it is possible to act and have an impact in places distant from our actual physical location. These information-based semantic environments are the infosphere (Floridi, 2002b; Van der Veer Martens, 2015; The Π Research Network, 2013). As a consequence of the information revolution and the inhabitation of the infosphere humans have become informational organisms or inforgs and they live an “onlife” – i.e. a life in between the online and offline:

“For example, a person driving using GPS is not clearly either online or offline. In these ways, categories developed in what we easily recognize as a new and special infosphere – the internet – are expanding naturally into the rest of our experience. Floridi calls this new status “onlife”(…). It is because of this that we are coming to see the whole world as the infosphere and ourselves as inforgs.” (The Π Research Network, 2013, p. 31).

¹⁸ For Floridi the critical investigations in (a) are done on information-theoretic grounds with the purpose of naturalization (Floridi, 2002b).

¹⁹ Similar to Williamson's (2000) ‘knowledge first’-approach in epistemology. Williamson argued that knowledge was the most fundamental concept, wherefore it could not be reduced and explained in terms of belief, justification, truth, etc. Instead, knowledge would be the concept through which an understanding and an account of belief etc. could be gained and developed.

According to Floridi, what enabled the information revolution was Alan Turing's basic idea of computation along with Claude Shannon's Mathematical Theory of Communication (Floridi, 2002a, 2002b; Van der Veer Martens, 2015; Adams & de Moraes, forthcoming; The Π Research Network, 2013). In Floridi's view the information revolution demanded the emergence of philosophy of information as an independent philosophical field.

In order to investigate and critically examine the information revolution, the infosphere, the inforgs, the onlife, and information itself as the most foundational concept, there is one important method employed within the Floridian conception of philosophy of information – The method of Levels of Abstractions (LoAs). A level of abstraction is the level through which something is observed, it is sets of typed variables over a given system (Floridi, 2005b; Van der Veer Martens, 2015; The Π Research Network, 2013). In other words, a LoA is a “model” or a “conceptual framework” through which the world is observed. “Through a LoA, an information agent (the observer) accesses a physical or conceptual environment, the system. LoAs are not necessarily hierarchical and they are comparable. They are interfaces that mediate the epistemic relation between the observed and the observer.” (Floridi, 2005b, section 3.2.2). For example, a programmer, a doctor, and a patient will observe a given medical database or system differently through their own level of abstraction as programmer, doctor, and patient, respectively. These different levels of abstractions employed by the three agents determine how they observe the database, what they observe, and why they observe it. The programmer's LoA might make him observe whether he can improve the database or the system by developing new and better algorithms. The doctor's LoA might make him observe the database or system in accordance to scientific and medical requirements – does the system support his medical practice in the best way? And the patient's LoA might make him observe the database or system in accordance to privacy issues – what kind of information about him does the database contain. Every agent can employ different LoAs at different times, as for example when the programmer himself is a patient. LoAs can also be built into systems through the typed variables. Floridi's example of a LoA within a system is a motion detector. The LoA within the detector is the set of typed variables ‘motion’ and ‘infrared radiation’, which define what the detector detects (Floridi, 2005b). Different LoAs have different outcomes and implications. Floridi's point in the adherence to the method of LoAs is that every philosophical or scientific investigation is undertaken through a specific LoA, whether or not the philosopher or scientist is aware of it. Therefore, the best critical investigations of specific questions or concepts are those where the LoAs are made explicit.

The method of levels of abstraction is essential to Floridi's conception of philosophy of information as it helps ask the right answerable open questions. For instance, absolute questions are questions which are asked outside any LoA, but as people can only approach a question through a LoA, absolute questions cannot be answered (The Π Research Network,

2013). The point is that the LoA within which a question is posed is the LoA within which the question must be answered. In other words, if a question is posed in one LoA and answered from another LoA, the answer – if any provided – will be unsatisfying and will create a mess. Therefore, the choice of LoA enables some questions and answers and excludes others. Or, as Floridi puts it: “Data as constraining affordances—answers waiting for the relevant questions—are transformed into factual information by being processed semantically at a given LoA (alternatively: the relevant question is associated to the right answer at a given LoA).” (2005b, section 3.2.2). Furthermore, the method of levels of abstraction is closely connected to Floridi’s constructionist²⁰ view of knowledge and information as well as his conception of truth. Thus, in his argumentation for why he does not endorse a correspondence theory of truth (cf. section 4.2.5), Floridi writes:

“Ultimately, LoAs construct models of data systems; they do not represent or photograph or portray or photocopy, or key-hole spy or map or show or uncover or fax or monitor or ... the intrinsic nature of the systems they analyze no more than an igloo describes the intrinsic nature of snow or the Parthenon indicates the real properties of stones. We neither discover nor invent the world; we design it.” (Floridi, 2004a, p. 663).

The method of level of abstractions, as well as the constructionist view of knowledge (i.e. that people construct knowledge rather than obtain it), ties into the notion of the infosphere:

“Finally, it is worth noting that Floridi distinguishes between what he calls a “minimalist” and “maximalist” interpretation of PI. Floridi claims that everything is information. The whole universe and everything in it is made of information – the universe is the infosphere. Broadly, this is because information is the broadest Level of Abstraction (...) at which to describe everything. (...) The maximalist approach to philosophy of information takes this idea seriously. PI is the philosophy of information design. As, according to this view, the world is made of information, this is a foundational philosophy. PI is the philosophy of everything.” (The Π Research Network, 2014, p. 33).

The conception of the universe as the infosphere follows from the concept of humans and systems as informational organisms – i.e. inforgs – which means that humans are explained in informational terms the same way as systems. Thus, in order to understand and explain the

²⁰ This term is special to Floridi and denotes a view of knowledge and information as constructed in an engineering sense (Floridi, 2004a, 2011a).

human brain, for instance, notions of ‘information processing’ and ‘computation’ are employed. Floridi (2002b) points out, that the goal is not to explain everything in terms of information just because it is possible, rather the goal is to investigate what it would be like for some concept (for example the human mind) not to be informational by nature. However, in the minimalist approach to PI, the infosphere is just the internet, including internet-based services (such as e-mail, social media, etc.), databases, and the like. In either case “[t]he infosphere is the environment constituted by the totality of information entities – including all agents – processes, their proprieties and mutual relations.” (Floridi, 1999, p. 44) or “the semantic environment in which millions of people spend their time nowadays” (Floridi, 2002b, p. 41).

1.3.2 Other interpretations of PI

Even though Floridi speaks of philosophy of information as a field, his conception of PI (cf. section 1.3.1) might be more of a project than a field.

In the article *An Illustrated Introduction to the Infosphere* (2015) Van der Vere Martens consequently speaks of “Floridi’s PI project” (p. 328), “his [Floridi’s] PI framework” (p. 320), or simply “Floridi’s PI” (p. 325). Thus, in her abstract, Van der Vere Martens writes: “This introduction to Luciano Floridi’s philosophy of information (PI) provides a short overview of Floridi’s work and its reception by the library and information studies (LIS) community, brief definitions of some important PI concepts, and illustrations of Floridi’s three suggested applications of PI to library and information studies.” (Van der Vere Martens, 2015, p. 317)²¹. This quotation points to the view that PI, as presented above (cf. section 1.3.1), including its key concepts (the information revolution, the infosphere, the inforgs, and the onlife) is Floridi’s project and conception of PI and not philosophy of information per se. This view is further supported by statements such as “Floridi’s (...) informational revolution” (Van der Vere Martens, p. 334), “Floridi’s reinvention of the term *infosphere*” (p. 332), “Floridi’s information macroethics” (p. 335), and “Floridi’s term *information organism*, or *inforg*” (p. 336). On the other hand, the many quotes of ‘Floridi’s such-and-such’ could also point to the conception that Floridi’s project is what constitutes philosophy of information per se and that the only way to conceive of this field and engage in it is Floridi’s way.

However, as seen in section 1.3 philosophy of information is a field which encompasses more than Floridi’s project and which has at least three different interpretations (i.e. ‘branch’,

²¹ It should be noted that Van der Vere Martens is sympathetic to Floridi’s project and sees it as a useful framework for library and information studies.

‘intersections’, and ‘world view’). To embrace a Floridian term, Floridi employs one LoA as interface to philosophy of information, with inforgs, infosphere, onlife, and information revolution as the typed variables. Others employ other LoAs as interfaces to philosophy of information with other typed variables – e.g. philosophy of information as various intersections, for instance between analytical philosophy and information studies (e.g. Fallis, 2009, 2014, 2015; Furner, 2004; Mai, 2013; and with emphasis on cognitive science and computer science Scarantino & Piccinini, 2010). As expressed by Adriaans (2012) in *Information on Stanford Encyclopedia of Philosophy*:

“Some authors like Floridi (...) present ‘Philosophy of Information’ as a completely new development with a capacity to revolutionize philosophy per se. Others (...) see it more as a technical discipline with deep roots in the history of philosophy and consequences for various disciplines like methodology, epistemology and ethics.” (Adriaans, 2012, introduction).

The ‘others’ that Adriaans refers to is an article by Lenski (2010) as well as the introduction to *Philosophy of Information* (2008) edited by Pieter Adriaans himself and Johan van Benthem (cf. the three interpretations in section 1.3). In addition to these ‘others’ Adams and de Moraes accept and defend Floridi’s idea of philosophy of information as an autonomous field but they do not accept “[t]he ideological aspect of Floridi’s proposal [which] is an ontological one: a replacement of a human-centric and conscious mind-centric world view with an information-centric view, in which “information” acquires the central role.” (Adams & de Moraes, forthcoming). *Philosophy of Information* (2008) is a Handbook within the series *The Philosophy of Science* published by Elsevier²². The handbook provides a range of different approaches to philosophy of information by writers from different fields, which include, but are not limited to, computer science, cognitive science, artificial intelligence, logic, epistemology, and philosophy of language – fields which are also closely connected to information studies. However, information studies is not represented in the Handbook as a field which contributes to philosophy of information. This stresses the conclusion Furner (2010) draws on the influence between philosophy and information studies as more or less one-directional from philosophy to information studies.

In the introduction to the handbook it is stated that it is an activist handbook in the sense that it “has a *performative* use, trying to create a new field by a ‘let it be.’” (Adriaans & van Benthem, 2008, p. 3). In their clarifications of the scope of the handbook and the new field it

²² This *Philosophy of Information* (2008) must not be confused with Floridi’s *The Philosophy of Information* (2011b).

tries to address, Adriaans and van Benthem point out that philosophy of information might mean two different things:

“‘Philosophy of information’ might mean philosophy of the information sciences, just as there is philosophy of the natural sciences, the life sciences, or humanities. Such methodological reflection on specific fields is absolutely necessary given the explosion of relevant technical research. (...) But there is also the parallel, and in some ways more ambitious aim of information as a major category of thought within philosophy itself, which might have the potential of transforming that whole field.” (Adriaans & van Benthem, 2008, p. 3).

These two ‘meanings’ are not directly compatible to the interpretations of philosophy of information as a field already addressed. The first ‘meaning’ bear resemblance to Floridi’s conception of philosophy of information as the foundation for library and information science, although Adriaans and van Benthem’s description of a ‘philosophy of the information sciences’ is broader. ‘The more ambitious aim’ encompasses both Floridi’s ‘world view’ conception of philosophy of information and philosophy of information as an interdisciplinary branch of philosophy.

1.3.3 PI as interdisciplinarity and intersections

Despite differences in terminology and key concepts through which to investigate ‘information’, philosophy of information in Floridi’s terms is in some way encompassed by both ‘meanings’ or ‘aims’ of the field as described by Adriaans and van Benthem: 1. as a philosophical foundation for the information sciences; 2. as the critical investigations of ‘information’ within philosophy itself (Adriaans & van Benthem, 2008; Floridi, 2002a, 2002b, 2004a; and Van der Vere Martens, 2015).

The present thesis endorses a slightly different interpretation of philosophy of information, that is philosophy of information as an interdisciplinary field comprised of the intersections between the contributing disciplines – e.g. computer science, cognitive science, artificial intelligence, (analytical) philosophy, information studies, and the like. This interpretation is not described in the body of literature which explicitly addresses philosophy of information – this literature is mostly comprised of Floridi’s work or work which refers to Floridi’s conception of philosophy of information (sometimes as if the two were synonymous). However, as Adams and de Moraes (forthcoming) point out, it is possible to conceive of philosophy of information as an interdisciplinary and autonomous field without the acceptance of Floridi’s world view and ontological commitments (e.g. the idea of the universe as the infosphere). As information studies is more or less neglected within philosophy (Furner, 2010) it is more fruitful to interpret philosophy of information as an interdisciplinary

field comprised of multiple intersections than to interpret it as an interdisciplinary branch of philosophy. The ‘intersection’-approach has the advantage that it enables contributions from within the other disciplines, which comprise philosophy of information, for example information studies. In this way contributions from information studies, cognitive science, computer science, artificial intelligence, (analytical) philosophy, etc. can all be instances of philosophy of information as long as they share the goal of philosophical and critical investigations of ‘information’ from their specific point of view. Thus, the intersection between information studies and analytical philosophy encompasses the philosophical and critical investigations of ‘information’ through the use of philosophical methods and frameworks²³ on ‘objects’ borrowed from information studies. Furthermore, although defined as various intersections philosophy of information can still provide the theoretical and philosophical foundation for the rest of information studies.

Philosophy of information as an interdisciplinary field comprised of intersections between computer science, artificial intelligence, cognitive science, information studies, (analytical) philosophy, etc. is the interpretation employed in this thesis²⁴. The explorations of the connections between information, misinformation, and disinformation will be conducted within a specific area of philosophy of information – that is, within the intersection of information studies and analytical philosophy represented by philosophy of language and epistemology. The ‘object’ is the notions of information, misinformation, and disinformation in the context of social network structures, automatic detection, and communication, the method is Carnap’s (1950) method of explication, and the framework is Gricean (cf. section 1.5 and Ch. 2).

1.4 The method of explication

The method of explication will be employed in this thesis as a method for the development of an understanding and evaluation of the concepts of information, misinformation, and disinformation and their interrelations (as philosophical concepts) with attention to the context of social network structures, communication, and automatic detection.

In *Logical Foundations of Probability* (1950) Rudolph Carnap presents and develops his method of explication which is a method to develop and evaluate definitions of a given concept within specific contexts. “By the procedure of *explication* we mean the transformation of an inexact, prescientific concept, the *explicandum*, into a new exact concept, the *explicatum*.”

²³ Most often from philosophy of language and epistemology.

²⁴ Except, of course, when Floridi’s notion of information is the object of analysis.

Although the explicandum cannot be given in exact terms, it should be made as clear as possible by informal explanations and examples.” (Carnap, 1950, p. 3). Thus, the method of explication consists of two steps:

- First the explicandum must be elucidated;
- then it must be transformed into an explicatum.

Besides being prescientific and inexact it is further specified that “[t]he explicandum may belong to everyday language or to a previous stage in the development of scientific language.” (Carnap, 1950, p. 3). Thus, in the case of the notions information, misinformation, and disinformation, the definitions and ordinary usages of the concepts as found in dictionaries serve as the explicanda (i.e. sections 1.2.1 and 1.2.2). This means that the descriptions and discussions of the dictionary-definitions of information, misinformation, and disinformation are the first step in an explication of information, misinformation, and disinformation. The second step is to provide more exact concepts of information, misinformation, and disinformation – i.e. the explicata. In other words, the relation between the two steps is that “[c]larifying the explicandum serves the purposes of specifying, perhaps in relatively crude terms, what is to be included and what is to be excluded. Once this has been accomplished we can meaningfully discuss possible explications of the term in question.” (Olsson, 2015, p. 61). The method of explication does not prescribe that the explicata (or possible explications) must be developed anew. It is perfectly fine to evaluate explicata that are already provided by other researchers or philosophers. Thus, possible explicata of information, misinformation, and disinformation will be found in philosophical accounts of these concepts, within the scope of philosophy of information as this field is where such accounts are offered.

Any explicatum has to be evaluated as “[a] concept must fulfil the following requirements in order to be an adequate explicatum for a given explicandum: (1) similarity to the explicandum, (2) exactness, (3) fruitfulness, (4) simplicity.” (Carnap, 1959, p. 5). The explicatum is evaluated within a specific context for its use which means that such a context has to be specified. This means that each philosophical definition (i.e. each explicatum) of information, misinformation, and disinformation, respectively, must be evaluated within a specified context. In this evaluation a distinction between two contexts can be made: the context set by the detecting-projects (that is the purposes of the detecting-projects as the context); and the context within which each philosophical account is developed (e.g. communication, learning, cognition, semanticization of the universe, etc.).

Hence, a more detailed description of the four evaluation criteria:

1. similarity between explicatum (i.e. a philosophical account of information, misinformation, or disinformation) and explicandum (i.e. definitions of information,

misinformation, or disinformation from dictionaries), which means that the explicatum must not be too distant from the explicandum. It is the explicandum (the pre-systematic concept) which determines what explicatum can be arrived at through the explication;

2. exactness in terms of formal formulations of explicatum within an established framework, i.e. the specified context (e.g. communication, learning, cognition, semanticization of the universe, etc.);
3. fruitfulness of the explicatum within the specified context, i.e. how fruitful is the explicatum as concept within that context (e.g. how useful is a specific philosophical notion of disinformation within the context of the detecting-projects or within the context of communication?);
4. and simplicity of the explicatum, i.e., in accordance with Occam's razor, the less complex an explicatum the better provided that the other requirements are met.

Simplicity is the less important criterion for evaluation as complex concepts can make perfectly good explicata as long as they satisfy the other evaluation criteria of similarity, exactness, and fruitfulness. As the evaluation criteria all come in degrees – in the sense that two explicata, for instance, both can be similar to a given explicandum although the one explicatum is more similar than the other – it is required that an explicatum fulfils the criteria “to a sufficient degree.” (Carnap, 1950, p. 7). Sufficiency and satisfaction are what is strived at in the evaluation of an explicatum, wherefore Carnap allows semiformal explicata and not just formal ones:

“Since the datum [the explicandum] is inexact, the problem itself is not stated in exact terms; and yet we are asked to give an exact solution. This is one of the puzzling peculiarities of explication. It follows that, if a solution for a problem of explication is proposed, we cannot decide in an exact way whether it is right or wrong. Strictly speaking, the question whether the proposed solution is right or wrong makes no good sense because there is no clear-cut answer. The question should rather be whether the proposed solution is satisfactory, whether it is more satisfactory than another one, and the like.” (Carnap, 1950, p. 4).

Thus, whether an account of information, misinformation, or disinformation as explicatum is satisfactory within the specific context where it is proposed as well as within the context of automatic-detection is what is evaluated in terms of fruitfulness, exactness, similarity, and simplicity. Every account of information, misinformation, or disinformation within philosophy of information is a candidate for an explicatum of one or more of these notions and there are especially many candidates for an explicatum of information.

1.5 The Inevitable

Information is a polymorph term (cf. section 1.2) and every field working with information has at least one notion or definition of the concept developed for more or less specific purposes without any connection to philosophy of information (cf. Adriaans, 2012; Cornelius, 2002; Floridi, 2008; Robinson & Bawden, 2014). Within philosophy of information multiple accounts of information have been developed and proposed either in the form of technical terms suited for specific uses or as overall terms intended to capture ‘the ordinary sense’ of information (for an overview see Adriaans, 2012; Adams & de Moraes, forthcoming; Capurro & Hjørland, 2003; Floridi, 2005b; Furner, 2010, 2014; and Robinson & Bawden, 2014). In philosophy of information the accounts of information are presented within different intersections or areas of the field, wherefore they do not necessarily figure²⁵ in the intersection of information studies and analytical philosophy (e.g. Hintikka, 2007; and most of the contributions to Adriaans & van Benthem, 2008). However, many accounts of information are present within this specific intersection either as explicit contributions to this part of philosophy of information or by reference, import, and use for different purposes (e.g. Budd, 2011; Dretske, 1981; Fallis, 2009, 2011, 2014, 2015; Floridi, 2004b, 2005a, 2005b, 2007, 2011b; Fetzer, 2004a; Fox, 1983; Scarantino & Piccinini, 2010; and Skyrms, 2010).

With misinformation and disinformation it is different. These notions are not polymorph in the same way as information and they have not received the same amount of attention – neither within philosophy of information nor outside the field. The few accounts of misinformation and disinformation offered within philosophy of information are all within the intersection between information studies and analytical philosophy (e.g. Fallis, 2009, 2011, 2014, 2015; Fetzer, 2004a, 2004b; Floridi, 2005b, 2011b; Fox, 1983; and Skyrms, 2010) and often they arise as a by-product of, or as brief additions to, an account of information (e.g. Fetzer, 2004a; Floridi, 2005b, 2011; and Skyrms, 2010).

The most prominent attempt to develop an account of misinformation is Fox’s (1983) conceptual analysis of the notion. The conceptual investigation is conducted in parallel with analyses of information and most of the conclusions drawn for misinformation are derived from conclusions drawn for information. Besides Godfrey-Smith’s (1989) account of the problem of error for informational semantics²⁶, Fox’s analysis seems to be the only regular

²⁵ However, this does not mean that they do not fit into this intersection. For example, Hintikka’s (2007) scope, object, and method – i.e. knowledge-seeking through questioning in database structures – fits well within this intersection.

²⁶ In the title referred to as misinformation.

attempt at an account of misinformation – the notion only earns mention or brief definitions in Dretske (1981, 1983, 2008), Fallis (2009, 2011, 2014, 2015), Fetzer (2004a, 2004b), Floridi (2004b, 2005a, 2005b, 2011b)²⁷, Scarantino and Piccinini (2010), and Skyrms (2010). In Fallis (2009, 2015) Fox is the reference in connection to misinformation and in Floridi (2005b) the definition of misinformation is a direct quote from Fox (1983). Moreover, in Dretske (1981) there is a definition of misrepresentation which can be analyzed as a determinant for misinformation and it is this account which Godfrey-Smith (1989) analyzes and criticizes under the headline ‘misinformation’.

Disinformation has earned more attention than misinformation, perhaps due to the aspect of deliberate deception (cf. section 1.2.1), which seems to generate the conception that disinformation is worse than misinformation (Fallis, 2015). Don Fallis is the philosopher and information studies scholar who has provided the most extensive, systematic, and conceptual analyses of disinformation. Fallis’ (2009, 2011, 2014, 2015) account(s) are based on the literature on lying, misleading, and deceiving and Grice’s work on cooperation (i.e. the Cooperative Principle and its maxims) and deception (from *Meaning*, 1957). Besides Fallis, Fetzer (2004a, 2004b) has offered accounts of disinformation. In Fetzer (2004a) the account of disinformation is more of a short description than an actual account and it is offered aside a short description of misinformation in a counterargument to Floridi’s (2004a) argument that false information is not information. Thus, the actual purpose for Fetzer (2004a) is to provide an account of information, where information can be false and thereby capture misinformation and disinformation. In Fetzer (2004b) a more extensive account of disinformation is offered divided into five different kinds. Floridi (2005b, 2011b) has also developed accounts of disinformation. However, these – together with accounts of misinformation – are mostly developed as by-products of, or brief additions to Floridi’s accounts of information, that is, as something which is not information. Furthermore, disinformation is mentioned in Dretske (1983) and Skyrms (2010).

All the actual accounts of misinformation and disinformation (i.e. the accounts where definitions of misinformation and disinformation are not mere by-products or brief additions to accounts of information) come with short definitions of information as well. Thus, Fallis (2009, 2011, 2014, 2015) and Fetzer (2004b) provide short definitions of information, and Godfrey-Smith (1989) challenges the account of misrepresentation within an account of

²⁷ Floridi (2007) is not mentioned here as this article does not mention or refer to the notion misinformation. However, the article deals with ‘false information’, which in Floridi (2004b) is used as definition for misinformation.

information. On the other hand, Budd (2011), Dretske (1981, 1983, 2008), Fetzer (2004a), Fox (1983), Floridi (2004b, 2005a, 2005b, 2007, 2011b), Scarantino and Piccinini (2010), and Skyrms (2010) have developed regular and extensive accounts of information, where misinformation (and sometimes disinformation) are at least mentioned. Add to these all the accounts that are solely of information and do not mention misinformation and disinformation, for example, Bar-Hillel and Carnap (1952), Hintikka (2007), Shannon (1948), and many of the contributions in Adriaans and van Benthem (2008).

Not all of these accounts can be fully analyzed, evaluated, and discussed as *explicata*. Given the scope of the present endeavor – that is, the investigation and exploration of the notions information, misinformation, and disinformation and how these are connected along with redefinitions that take these connections into account (i.e. possibly new *explicata*) – accounts developed by a few researchers and philosophers offer themselves as inevitable. These are the accounts by Dretske (1981, 1983, 2008), Fallis (2009, 2011, 2014, 2015), Floridi (2004b, 2005a, 2005b, 2007, 2011b), and Fox (1983). These four philosophers and researchers are some of the most influential within philosophy of information when it comes to the notions of information (Dretske, Floridi, and to some extent Fox), misinformation (Fox), and disinformation (Fallis). Furthermore, they are interconnected through extensive references.

1.5.1 Dretske and Floridi

Philosopher and logician Luciano Floridi's concept of information must be taken into account although it is developed within a conception of philosophy of information different from the one endorsed in this thesis. Floridi's conception of information is the most dominant conception within philosophy of information and it serves as the starting point for, or the centerpiece of, most discussion within the field (e.g. Fallis, 2011; Fetzer, 2004a; Furner, 2010; Robinson & Bawden, 2014; Scarantino & Piccinini, 2010; The Π Research Network, 2013; and Van der Veer Martens, 2015). Furthermore, Floridi provides definitions of misinformation and disinformation (although not to the same extent as information) which enables exploration of the connections between the three notions within Floridi's own accounts. Floridi's definition of misinformation is quoted from Fox (1983) and his notion of semantic information is partly inspired by Grice and Dretske.

"The philosopher Fred Dretske [1932-2013] was a central figure in the (re-) establishment of interest in the concept of information in current Anglo-American analytic philosophy." (The Π Research Network, 2013, p. 18) and his account of semantic information has proved rather influential within philosophy of information (e.g. Adams & de Moraes, forthcoming; Budd, 2011; Capurro & Hjørland, 2003; Fallis, 2009; Floridi, 2005a, 2005b, 2011b; Furner, 2010; Godfrey-Smith, 1989; Mai, 2013; Robinson & Bawden, 2014; Scarantino & Piccinini, 2010; and

The Π Research Network, 2013). Dretske's account of information entails an account of the possibility of misrepresentation as determinant for misinformation.

Based on the Mathematical Theory of Communication²⁸ by Shannon (1948) (cf. Dretske, 1981, p. 237, note 1), Dretske develops a semantic theory of information with the purpose of developing an information-theoretic account of knowledge (Dretske, 1981). Dretske "was not the first analytic philosopher to offer a systematic treatment of information. (...) in the 1950s, Rudolf Carnap (1891-1970), together with the Israeli mathematician and linguist Yehoshua Bar-Hillel (1915-1975), attempted to develop an account of semantic information based on Shannon's theory." (The Π Research Network, 2013, p. 18). However, due to the ordinary language turn in analytic philosophy in the late 40's, Bar-Hillel and Carnap's account was more or less ignored. Dretske's *Knowledge and the Flow of Information* (1981) came with a new turn in analytic philosophy, where the interest in formal modeling was renewed (The Π Research Network, 2013). Dretske sets out to develop an account of semantic information, that is, an account of the ordinary concept of information. Although Dretske's theory is developed with the outset of Shannon's theory, the ordinary concept, which Dretske wants to capture, is the concept of information connected to knowledge, news, learning, and intelligence – and not the mathematical and information-theoretic concept (cf. section 1.2.2).

Dretske begins his endeavor with

"some of the fundamental ideas of communication theory. (...) *not* because this theory (in its standard interpretation and application) tells us what information is. It does not. It does not even try. Rather, I begin here because the underlying structure of this theory, when suitably supplemented, can be adapted to formulate a genuinely *semantic* theory of information, a theory that can be used in cognitive and semantic studies." (Dretske, 1981, pp. ix-x).

In Dretske's theory of semantic information, information is an objective entity. Information is objective in the sense that it exists in the world independent of agents and it is defined by truth. If it is not true it is not information. Information is objective because it consists in signals, signals that carry information that some state of affairs or condition has obtained (i.e. is the case), and they do so independently of whether or not any agents get or understand these signals. Signals carrying information about the information at a source is the legacy from Shannon's theory. Dretske's addition is the semantic aspect, the informational content –

²⁸ By Dretske referred to as 'The Mathematical Theory of Information' or 'Communication Theory' (Dretske, 1981, p. 3 and p. 237, note 1).

not that the signal carries information or how much information it carries, but which information the signal carries. This addition leads to Dretske's definition of semantic information in terms of informational content:

INF_{DR} "*Informational content*: A signal r carries the information that s is F = The conditional probability of s 's being F , given r (and k), is 1 (but, given k alone, less than 1)" (Dretske, 1981, p. 65).

Dretske specifies that a signal r "can be any event, condition, or state of affairs the existence (occurrence) of which may depend on s 's being F ." (Dretske, 1981, p. 65). This means that a signal (e.g. the perception of a flower) can only carry the information that s is F (e.g. that the flower is yellow) if it is true that the flower is yellow (i.e. the probability is raised to 1) and if the flower's being yellow is not provided by the agent's background knowledge, i.e. what he already knows (k). According to Dretske (1981), this definition satisfies the three conditions for information, which he has specified:

- "(A) The signal carries as much information about s as would be generated by s 's being F ." (p. 63).
- "(B) s is F ." (p. 64).
- "(C) The quantity of information the signal carries about s is (or includes) that quantity generated by s 's being F (and not, say, s 's being G)." (p. 64).

This means that misinformation, defined as that which is false, is not a kind of information, it is something else entirely (Dretske, 1981, 1983, 2008). To use the famous example by Dretske himself "..., *false* information and *mis*-information are not kinds of information – any more than decoy ducks and rubber ducks are kinds of ducks." (1981, p. 45). This is as far as Dretske goes when it comes to misinformation. Instead he focuses on misrepresentation, as the source of falsity. Misrepresentation occurs when the informational link between an information source and a type (e.g. a semantic structure) is not preserved in a token (e.g. a belief) of that type, which means that the belief can be false (Dretske, 1981). Thereby misrepresentation may lead to misinformation or disinformation, which is added to the example in the '*Précis of Knowledge and the Flow of Information*' (Dretske, 1983, p. 57). Apart from the explanation in terms of misrepresentation, misinformation is not defined, however the reference (the use) points to the definitions from the dictionaries (cf. section 1.2.1), except for the fact that misinformation is not wrong, bad, or false information. Misinformation is not a kind of information, since falsehoods cannot be information (Dretske, 1981, 1983, 2008).

The notion of information developed by Dretske is, as mentioned, a notion of semantic information. Semantic information is the type of information most often at play in philosophy

of information and it is this kind of information that misinformation and disinformation often are contrasted to. According to Floridi (2005a) there is no consensus on the definition of semantic information²⁹. The Standard Definition of Information (SDI)³⁰, which stipulates that information is well-formed, meaningful data, is still debated (Floridi, 2005a).

Floridi enters this debate with Dretske and Grice in hand arguing that SDI needs revision by the addition of a necessary truth-condition. This revision is needed because, according to Floridi, declarative, objective, and semantic information (DOS-information, which is the concept of information Floridi is after³¹) encapsulates truthfulness, whereas SDI presumes that the truth-value (the alethic value) is an extraneous subsequent addition (Floridi, 2005a) wherefore it cannot be the standard definition within a DOS-conception of information. By the addition of the truth-condition to SDI Floridi arrives at RSDI (Revised Standard Definition of Information) in which information is defined as

INF_{FL(2005a)} “well-formed, meaningful,³² and truthful data” (Floridi, 2005a, p. 366)

(the thesis that information requires truth in order to be information is known as The Veridicality Thesis). Floridi’s formulation of RSDI as a definition of information within the DOS-conception is objective in the sense that it is independent of agents or at least independent of informees. Hence, information can be present even though agents cannot access it (Floridi, 2005a). Because Floridi’s work is mostly concerned with the notion of information and its veridical nature, not much is said about misinformation and disinformation³³, which, according to Floridi, are not kinds of information as they are false.

²⁹ However, the requirements for objectivity and truthfulness seem to be widely agreed upon in philosophy of information, and are mostly questioned from outside the field.

³⁰ In Floridi (2005b, section 1.2) the standard definition is named The General Definition of Information (GDI) however the definition is the same i.e. well-formed, meaningful data. In Floridi (2003) it is the abbreviation GDI which is in use for the General Definition of Information. Here the revised definition, which entails truth, is called SDI for Special Definition of Information and it is identical to RSDI from Floridi (2005a). Thus, in Floridi (2003) SDI stand for Special Definition of Information and reads well-formed, meaningful, and truthful data, whereas in Floridi (2005a) SDI stands for Standard Definition of Information and reads well-formed, meaningful data. In Floridi (2011b) the abbreviations GDI, for well-formed, meaningful data, and GDI*, for well-formed, meaningful, and truthful data, are used. In Ch. 8 in Floridi (2011b) GDI* is summarized and consequently referred to as [SI] for Semantic Information (cf. section 4.2.1).

³¹ Floridi (2005b) acknowledges that there can be other forms of information which are not declarative (instructional information) and not semantic (environmental information).

³² ‘Meaningful’ refers to the semantics. Floridi (2005a) speaks of meaningfulness independent of informees (pp. 358-359), whereas Dretske speaks of meaningfulness independent of agents (Floridi, 2005a, p. 359).

³³ Both notions are mentioned in parsing – misinformation with a reference to Fox (Floridi, 2005b).

Floridi (e.g. 2005a, 2005b) follows Dretske and Grice – in his argumentation that false semantic content is not information – by reference to Dretske’s decoy duck-example as well as Grice’s (1989) statement that “[f]alse information is not an inferior kind of information; it just is not information.”³⁴ (Grice, 1989, p. 371). However, Floridi (2011b) formulates his own definitions of misinformation and disinformation stating that

- | | |
|--------------------------|--|
| MIS _{FL(2011b)} | “misinformation is ‘well-formed and meaningful data (i.e. semantic content) that is false’.” (Floridi, 2011b, p. 260). |
| DIS _{FL(2011b)} | “‘Disinformation’ is simply misinformation purposefully conveyed to mislead the receiver into believing that it is information.” (Floridi, 2011b, p. 260). |

Floridi’s (2011b) definitions of misinformation and disinformation resemble the definitions from the dictionaries (cf. section 1.2.1). The only difference is that Floridi uses the formulation ‘semantic content’ instead of ‘information’ as information on his account requires truth and therefore cannot be part of the definition of something false. On Floridi’s accounts, as well as Dretske’s, misinformation and disinformation have not been developed in their own terms the same way as semantic information. Semantic information, as well-formed, meaningful, veridical data (Floridi, 2005a, 2005b) or as an objective and true entity devoid of agents (Dretske, 1981, 2008) do not resemble the definitions from the dictionaries (i.e. the explicanda; cf. section 1.2.2). The dictionaries define information as knowledge learned and/or communicated, intelligence, news, and education – all concepts which require agents. Information as computation or mathematically defined quantity (information theory) does share some resemblance with the outset of Dretske’s and Floridi’s notions. However, both Dretske (1981) and Floridi (2005a, 2005b) emphasize that their notions of information are semantic notions. Dretske’s notion of information is developed on information-theoretic grounds by the addition of aboutness – i.e. what the information is about. Floridi’s notion of information is prior to computation as “[philosophy of information] analyses the latter as presupposing the former.” (Floridi, 2002b, p. 138). For Floridi and Dretske, the differences between information, misinformation, and disinformation are cast in terms of truth and falsity. Hence, Dretske and Floridi adhere to a true/false dichotomy in regard to information (as true) vs. misinformation and disinformation (as false).

³⁴ This point is stated in Grice (1989) in relation to the Quality Maxim without any argumentation.

The adherence to a truth/false dichotomy and the consequences it has for the possible connections between information, misinformation, and disinformation is one of the main differences between Floridi and Dretske, on one hand, and Fox and Fallis, on the other hand.

1.5.2 Fox and Fallis

Researcher Christopher Fox's book *Information and Misinformation* (1983) is more or less the only account of misinformation (besides Dretske's account of misrepresentation). Fox provides a conceptual analysis of utterances that contain the notions information and misinformation. The analysis is conducted through various tests, some of which are based on Grice (1989), and the conclusion with regard to misinformation has been widely accepted as a standard definition of misinformation within philosophy of information. Furthermore, Fox's notion of information is not defined by truth and it is often referred to in discussions of whether or not information can be alethically neutral³⁵ (The Π Research Network, 2013).

Philosopher and information studies scholar Don Fallis has provided the most extensive (and most recent) conceptual analyses and accounts of disinformation. The accounts have continuously developed and changed from Fallis' earliest definition in 2009 to his latest definitions in 2014 and 2015. In addition to the overall changes in how disinformation is defined, the various accounts have different scopes and outsets. Thus, Fallis' (2011) account is developed from a critique of Floridi's proposals to a definition of disinformation; Fallis' (2009 and 2015) accounts provide critiques of Fetzer's (2004a and 2004b) accounts; and Fallis' (2014) account is, among others, based on Grice (1989). Within all his accounts of disinformation Fallis provides definitions of information and misinformation as well.

When Fallis (2011) points to the accounts of misinformation and disinformation given by Floridi, he does so with the purpose of developing and formulating a conceptual account of disinformation. The purpose is to encapsulate all the intuitive senses of disinformation, for example true disinformation. In *The Varieties of Disinformation* (2014) Fallis suggests that

DIS_{FA(2014)} "disinformation is information that is *intentionally misleading*. That is, it is information that – just as the source of the information intended – is likely to cause people to hold false beliefs." (Fallis, 2014, p.137)

³⁵ Alethic Neutrality = that alethic values (i.e. truth-value such as truth or falsity, or no alethic value) are not embedded in, but supervene on semantic information. Alethic neutrality implies that well-formed, meaningful data qualify as information independent of truth-value (Floridi, 2005a).

Information is defined as

INF_{FA(2014)} “something that has *representational content*.” (Fallis, 2014, p. 137)

regardless of its truth-value. This means that information, on Fallis’ account, is not defined by truth. The distinction between intention and intentionality (cf. sections 5.1-5.3) is important in Fallis’ analysis because it enables him to capture cases of so called “*side effect disinformation*” (Fallis, 2014, p. 138) as genuine cases of disinformation. These cases arise when the misleadingness is not intended but it is intentional because it is foreseeable³⁶. The requirement for intentional misleading is an addition to, and specification of, the account of disinformation, which Fallis offers in *What is Disinformation* (2015³⁷). In this account

DIS_{FA(2015)} “disinformation is misleading information that has the *function* of misleading someone.” (Fallis, 2015, p. 413).

More precisely it is the representational content of information that has the function to mislead (Fallis, 2015). A function can be acquired in one of two different ways, either by evolution or by design: “For instance, a heart has the function of pumping blood because that is what it *evolved* to do; by contrast, a chair has the function of being sat on because that is what it was *designed* to do. (...) In other words, the designer of the artifact *intended* it to have that function.” (Fallis, 2015, p. 413). Thus, ‘intention’ is built into the notion of function through the notion of design. According to Fallis (2015) disinformation’s function to mislead can be of either kind, however, most prototypical instances of disinformation such as lies and propaganda are misleading because they are intended and designed to be. The distinction between intention and intentionality, which is added in *The Varieties of Disinformation* (2014), is one of three features which distinguish disinformation from mere lying. Cf. section 1.1.1, to lie is to make a believed-false statement that is intended to mislead³⁸. Thus, in this respect disinformation is broader than lying, although lying is still a kind of disinformation. The other two features are that disinformation does not have to be a statement and it does not have to be false. That disinformation does not have to be a statement allows the inclusion of

³⁶ Fallis’ example of *side effect disinformation* is researchers who insert inaccurate information into Wikipedia to see whether it gets corrected and how long it takes (Fallis, 2014). These researchers do not intend to mislead anybody but because it is foreseeable that the inaccurate information might mislead some users before it gets corrected, the misleadingness is intentional.

³⁷ Fallis’ notion of disinformation develops throughout his publications. In this connection it is important to note that ‘*What is Disinformation?*’ (2015) was actually written before ‘*The Varieties of Disinformation*’ (2014). The chronology of the articles has been reversed due to different times of publication. Fallis points out that his official view is that disinformation has the function to mislead (Fallis, 2015, personal communication).

³⁸ With the exception of ‘bald-faced lies’.

visual disinformation (Fallis, 2014). Visual disinformation is based on the possibility of non-verbal or gestural misleading and deception (cf. Mahon, 2008; and section 1.1). That disinformation does not have to be false enables the inclusion of *true disinformation* (Fallis, 2014). The possibility of true disinformation is derived from the definitions of misleading and deceiving based on Grice's notion of false implicature. The true disinformation arises when what is literally said is true but what is implicated is false, whereby the conclusion drawn from the implicature is likely to mislead (Fallis, 2014; cf. sections 1.1.2 and 2.3.2-2.3.3; and Grice, 1967).

In addition to the definition of information (as representational content regardless of its truth-value) and the definition of disinformation (as intentionally misleading information) Fallis has a definition of misinformation:

MIS_{FA}(2015) "*misinformation* is information that is inaccurate *and* misleading."
(Fallis, 2015, p. 423, note 7).

As with both information and disinformation, Fallis does not require a specific truth-value for misinformation, i.e. he does not require that it is false only that it is inaccurate. Information, misinformation, and disinformation are connected in a way which resembles the picture from Vocabulary.com (cf. section 1.2.1):

- Information is that which has representational content,
- misinformation is that subset of information that is inaccurate in one way or another and thereby misleading, and
- disinformation is that subset of misinformation, which is intentionally misleading.

Information researcher Christopher John Fox (1983) precedes Fallis on some of these points concerning the truth-value of information and misinformation. In philosophy of information Fox is often referred to as the main figure arguing for alethic neutrality for information, i.e. that information can be either true or false (Floridi, 2005a, 2005b; and The II Research Network, 2013) and he seems to be the sole researcher who have tried to develop an account of misinformation. The accounts of misinformation and information are developed simultaneously with the main emphasis on information. Fox has a reductionist approach to the analysis of information (and misinformation) where he analyzes information in terms of propositions – i.e. he reduces information (and misinformation) to propositions. Following his extensive analyses of information, misinformation, inform, and misinform based on questions of entailment Fox concludes that:

"All the evidence is in favor of the thesis that misinformation must be false. Since information may be false, we see that misinformation is a species of information, just as misinforming is a species of informing. Thus a coherent picture of the relationships

between informing, information, misinforming, misinformation, and truth has emerged: informing does not require truth, and information need not be true; but misinforming requires falsehood, and misinformation must be false.” (Fox, 1983, p. 193).

Disinformation and disinforming are not parts of the account developed by Fox. If Fallis’ definition is accepted and disinformation exists in a true variety, then disinformation will not easily become a part of Fox’s account. When misinformation in Fox’s terms requires falsity (instead of inaccuracy or unintended misleading, as Fallis would put it) it cannot capture, as a subset, something which can be true. Therefore, disinformation in Fallis’ terms cannot be a subset or a species of misinformation in Fox’s terms.

Misinformation and disinformation are developed as philosophical notions in their own rights in the accounts given by Fox and Fallis, respectively. For Fox misinformation is a kind of information, namely that kind which is false, which means that information can be both true and false, i.e. information is alethically neutral. In Fallis’ account disinformation is defined as intentionally misleading information and it is a subset of misinformation. Misinformation is defined as inaccurate and thereby misleading information and is a subset of information. Information is defined as that which has representational content regardless of truth-value, whereby it is also alethically neutral.

As already stated, Dretske, Floridi, Fallis, and Fox and their various accounts of information, misinformation, and disinformation are inevitable in the exploration of these three notions and their connections within philosophy of information. Dretske and Floridi are grouped together as their accounts have a main emphasis on information – they are primarily accounts of information. Moreover, they both require that information is inherently true. Fallis and Fox are grouped together as their accounts are of disinformation and misinformation, respectively. Although Fox’s account is also, and perhaps primarily, an account of information, it is the single most influential account of misinformation provided within philosophy of information. Furthermore, both Fox and Fallis argue that information is alethically neutral, i.e. that information is not inherently truthful.

1.6 Three asymmetries emerge

Through the examination of the different accounts of information, misinformation, and disinformation, both the accounts in dictionaries and especially the philosophical accounts, it becomes clear that there are three asymmetries between the notions of information, misinformation, and disinformation. The first asymmetry is in terms of meaning, the second is in terms of truth, and the third is in terms of intentions. The ‘meaning’-asymmetry concerns what is meant by meaning in accordance to the various definitions – i.e. is it strictly semantic

meaning (Dretske and Floridi) or is pragmatic meaning present as well (Fox and Fallis)? The 'truth'-asymmetry concerns the differences in the definitions in accordance to truth and falsity – i.e. is information inherently truthful (Dretske and Floridi) or is it alethically neutral (Fox and Fallis)? And what about misinformation and disinformation do they require falsity (Fox, Dretske, Floridi)? The 'intention'-asymmetry concerns the relations between the notions and active agents as specified by the definitions – i.e. are the notions defined as agent-dependent (Fox and Fallis) or as independent of active agents (Dretske)?

The asymmetries derive from the fact that, although some accounts contain all three notions, the notions of information, misinformation, and disinformation are not fully discussed and defined in relation to one another. That is, their connections have not been explored systematically. Most often (cf. sections 1.5.1-1.5.2) one (maximum two) of the notions are in focus as the definiendum, with the other two (often misinformation and disinformation) referred to in a sense which resembles the accounts from the dictionaries. Or, if misinformation or disinformation is the notion in focus, information is referred to with a note on whether or not it requires truth – i.e. whether or not the Veridicality Thesis is endorsed.

To summarize the central points from the philosophical accounts examined so far, misinformation is defined as unintended misleading information or simply just false information (Fox, 1983; Floridi, 2005b; dictionaries, cf. section 1.2.1) whereas disinformation is defined as intended or intentional misleading information (Fallis, 2009, 2015; and dictionaries, cf. section 1.2.1). Both unintended and intended/intentional misleadingness need an agent to intend or un-intend the misleadingness because intentions cannot exist or be present without agents. This means that the definitions of misinformation and disinformation are agent-relative, or dependent on active agents; it takes an agent to intend to mislead – i.e. to disseminate disinformation – and thereby to disinform and it takes an agent to un-intend to mislead – i.e. disseminate misinformation – and thereby to misinform. Furthermore, as the misleadingness or the falseness of misinformation is unintended there must be another intention present (or no intention present) for the agent when he disseminates the misinformation. The intention present might be the intention to inform or provide someone with information (cf. sections 5.3-5.4). Even though 'intention' and active agents are not inherent in the notion of information, to inform is an action and actions require active agents to perform them. An agent can intend to inform or provide someone with information, but that intention is not part of the definition of information the same way as the intention to mislead or the un-intention to mislead are parts of the definitions of disinformation and

misinformation³⁹, respectively. In the case of misinformation, described above, the intended provision of information becomes a provision of misinformation due to the agent's unawareness or ignorance towards the accuracy or truth-value of the proposition in question. The intention of non-misleadingness is not inherent in Dretske's and Floridi's notions of semantic information as this notion is defined as inherently truthful instead. Misinformation, and in Floridi's case also disinformation, are defined as false and not a kind of information. However, this leaves out the possibility of true disinformation which Fallis advocates. When disinformation is defined in terms of 'misleading' and 'deceiving' based on false implicatures it follows that something more than strictly semantic meaning is present. Disinformation exhibits some pragmatic features as well when it comes to meaning. Pragmatic features of meaning are present in Dretske's account of misrepresentation but not as part of the definition and when Floridi speaks of 'meaning' it is in a strictly semantic sense. The pragmatic vs. semantic features of misinformation and disinformation ties into the connection between the definitions and active agents. If a definition requires pragmatic features of 'meaning' it must be agent-dependent. Whether truth and falsity are satisfactory as requirements for information and misinformation/disinformation, respectively also depends on whether only semantic meaning is a part of the definition or whether pragmatic meaning is present as well. When pragmatic meaning and false implicatures are allowed true disinformation becomes a possibility.

To return to the active agents, when it comes to information, or more precisely semantic information, the picture seems more complex. As stated above, it takes an agent to inform because to inform is an action, but active agents and intentions are not part of the definition of semantic information defined as well-formed, meaningful, and veridical data (Adriaans, 2012; Floridi, 2005b, 2007, 2011b, etc.; The Π Research Network, 2013). Information is defined as objective and truthful independent of agents, which first become part of the picture when it comes to belief- and concept formation and knowledge (Dretske, 1981) or the act of informing. This means that agents and their intentions are first considered relevant when it comes to the use of information. However, some use of information might constitute acts of misinforming or disinforming or even turn the truthful information into disinformation (cf. Fallis, 2014). When the notions of information, misinformation, and disinformation are defined as alethically neutral (Fox, 1983; and Fallis, 2009, 2011, 2014, 2015) there is no conflict in the shift of status between information, misinformation, and disinformation, for the same utterance. The status depends on the meaning of the utterance and whether or not it is

³⁹ If misinformation is defined as unintended misleading – and not as false information.

misleading, either intended or unintended. When the definitions of information, misinformation, and disinformation come with fixed truth-values, these shifts are not possible and their explanations must be found elsewhere. The ‘truth’-asymmetry – i.e. the differences in requirements for truth, falsity, and alethic neutrality – is not dependent on whether or not a definition is agent-relative. It is possible to require specific truth-values for notions that are defined agent-dependently (misinformation in Fox, 1983 and Dretske, 1981, and disinformation in Floridi, 2011b) and for notions that are defined independent of agents (information in Dretske, 1981 and Floridi, 2011b).

Due to their definition-bound relation to intentions in some way or other, the notions of misinformation and disinformation can never be objective notions existing independently of agents the same way as information is claimed to be (in Dretske, 1981 and Floridi, 2005a). Misinformation and disinformation therefore must be context-dependent notions where the application to given propositions differ from agent to agent and from context to context, which is the reason why the spatio-temporal dimensions are considered crucial in the PHEME-project⁴⁰. Thus, misinformation and disinformation are created by agents either willingly (disinformation) or unwillingly (misinformation). This conclusion even holds for misinformation in Dretske’s terms – i.e. that which is false and therefore is not information – because misrepresentation, as the source of the false belief, which constitutes the misinformation, occurs within or by the agent (Dretske, 1981), it is an agent who misrepresents the information and thereby risks to generate falsities (misinformation) or does so purposefully (disinformation).

All these observations are preliminary and merely suggest various implications which arise from the asymmetries. In order to pursue the implications of the asymmetries between the three notions information, misinformation, and disinformation these three notions must be analyzed, discussed, and defined in connection to one another in terms of their relations to meaning, truth, and intentions. They must be evaluated as explicata in accordance to one another within the contexts where they are proposed and within the context of automatic detection due to the scope of the detecting-projects.

The detecting-projects deal with written language in social network structures and their aim is automatic-detection of misinformation and disinformation within short semantic

⁴⁰ In the PHEME-project the spatio-temporal dimensions are also crucial for the notion of information. Information is bound by its spatio-temporal origin. In the original spatio-temporal context the information might be true, whereas in another spatio-temporal context the same piece of information might be false (Derczynski & Bontcheva, 2014a). Thus the notion of information used in the PHEME-project is one of alethic neutrality.

structures. In the PHEME-project (2014) emphasis on the author and the spatio-temporal dimensions – i.e. the context – of specific tweets or Facebook-updates are the key to determine whether the tweet or Facebook-update is speculation, controversy, misinformation, or disinformation. In Kumar and Geethakumari's (2014) algorithm-project the original author of a given tweet and his/hers credibility and trustworthiness is the determinant of whether a tweet is labelled information or misinformation. Taken together, the detecting-projects have their main emphasis on language use in context and the credibility and intentions of authors when they communicate through online social network structures. Therefore, Grice's cooperative perspective on communication – with his distinctions of word-/sentence-meaning and speaker-meaning, as well as his theory of meaning in terms of speaker-intentions – lends itself as a fruitful framework for the exploration, understanding, and explication of information, misinformation, and disinformation.

Ch. 2 Theoretical Framework

2 Grice – Communication as cooperation

Paul Grice (née Herbert Paul Grice, 1913-1988) was a philosopher of language. Although his work is included under the sub-field ‘ordinary language philosophy,’ he had an eye for formalism as well as general theories and principles. Grice is most famously known for his theory of communication as a cooperative enterprise, as well as his insight that a distinction must be made between context-invariant meaning⁴¹ and context-dependent meaning⁴², determined by speakers and their intentions. Grice’s theories of meaning and cooperation concern verbal and gestural communication and are highly influential in the literature on lying, misleading, and deceiving – a literature which is central in the distinctions between information, misinformation, and disinformation – and have also had (and still have) a great influence on the notion of information in philosophy of information (and epistemology) (cf. Ch. 1). Grice’s recognition of the speakers’ intentions as fundamental to meaning and communication renders his theories fruitful as a framework for the exploration of the connections and differences between information, misinformation, and disinformation. More precisely, Grice’s theory of meaning and the Cooperative Principle incl. its maxims will in this dissertation serve as the framework for discussions of the emerged asymmetries in terms of meaning, truth, and intention.

⁴¹ Word- and sentence-meaning.

⁴² Utterer’s (speaker-) meaning.

2.1 Ordinary language philosophy

“The other day a philosopher of science in a university quite a long way from Oxford asked me whether I thought that “The Ordinary Language Approach to Philosophy” had anything to contribute to the Philosophy of Science. Finding this question difficult to handle for more than one reason, I eventually asked him what he meant by “The Ordinary Language Approach to Philosophy.” He replied that he had been hoping that I would not ask that question, as he did not know much about the matter. Perhaps he thought that the reference of this phrase ought to have been immediately clear to me, since it was intended merely to pick out the sort of philosophizing in which I myself (and others at Oxford) habitually engage. Unfortunately I do not find it by any means easy to give a general characterization of the philosophizing in which I engage; indeed I am not sure that it is all of one sort; moreover, I am sure that one could find numerous methodological divergences among Oxford philosophers, though there does, no doubt, also exist a noticeable family resemblance.” (Grice, 1958, p. 171).

Ordinary language philosophy (sometimes known as ‘Oxford philosophy’ or ‘linguistic philosophy’) emerged out of the analytic tradition and flourished in Oxford after the Second World War. The ‘family resemblance’, which Grice refers to, is the belief that analyses of ordinary language and its use are fruitful for philosophy. Despite different suggestions as to who is the actual originator of ordinary language philosophy it was “under Ryle’s and Austin’s guidance it developed in directions not envisaged, and in some cases explicitly not approved, by its putative mentors.” (Chapman, 2005, p. 34). According to Chapman (2005) philosophers such as Wittgenstein, Moore, Russell, and Frege have been proposed as the actual originator of ordinary language philosophy. However, their influence might be of a more implicit kind through their great influence on the analytic tradition – they represent the formalist approach to language, where ordinary language is imperfect and needs improvement before it is suitable for philosophical purposes (Chapman, 2005). In contrast “Austin plainly viewed ordinary language as a wonderfully subtle and well-contrived instrument, one which is fashioned not for idle display but for serious (and nonserious) use.” (Grice, 1989, p. 384). In this way Austin has an informalist approach to language, where ordinary language and language use in itself is suitable for philosophical inquiry. Austin believed that ordinary language could provide insights to philosophical concepts through the way people talk about experience (Chapman, 2005).

Grice is often considered as one of the ordinary language philosophers: he was part of the discussion group at Oxford initiated by Austin in the late 1940’s and he is concerned with language use as an important component in philosophical analysis (Grice, 1989). However, Grice was “convinced of the necessity for explanations that took the form of general theories”

(Chapman, 2005, p. 62) and he became increasingly unsatisfied with Austin's approach to language and the idea that language use is sufficient for philosophical inquiry and understanding (Chapman, 2005). Furthermore, "Grice also differed from Austin in his growing conviction that a distinction needed to be recognized between what words literally or conventionally mean, and what speakers may use them to mean on particular occasions." (Chapman, 2005, p. 62). This distinction between word-meaning and utterer's meaning is one of Grice's main contributions to philosophy of language, as well as linguistics, and it ties into the distinction between semantics and pragmatics.

2.1.1 Semantics and pragmatics

Austin's key insight and main contribution to the philosophy of language was his William James Lectures, delivered at Harvard University in 1955 and published posthumously in 1962 as *How to Do Things With Words*. The idea developed throughout these lectures is that to speak is to act. Thus, when people say something they are performing specific speech acts. Austin starts his lectures with a distinction – a dichotomy – between performatives and constatives. Performatives are utterances which, through the use of performative verbs, count as that action, which the verb denotes. For example, when someone says "I bet you 100 pounds" the mere expression of the verb 'bet' in this sentence is the act of betting itself. Constatives on the other hand, are statements without performative verbs, wherefore the utterance is not an action. During the lectures Austin realizes that constatives are actions themselves but of another kind than performatives. He therefore cancels the dichotomy and introduces 'illocutionary act' instead as a label for a variety of different classes of speech acts. Illocutionary acts – the act 'in saying' – is the second step in a three step 'hierarchy' of speech acts, which are all present in a given utterance:

- Locutionary act – the act 'of saying', i.e. the mere expression of specific words and sentences with a specific meaning (in Frege's terms of 'sense' and 'reference').
- Illocutionary act – the act 'in saying', i.e. the act performed in expressing the locution, e.g. stating, warning, answering, questioning, describing, making an appointment, etc.
- Perlocutionary act – the act 'by saying', i.e. the response, feelings, consequences, etc. which the speaker wants to produce in his audience.

In this way the locution of the utterance

"Close the door"

is close the door, meaning by 'close' close, and by 'the door' *the door* (i.e. a specific door), the illocution is an order or a request to close the door, and the perlocution is persuasion to close the door (i.e. to produce the consequence of the door being closed) (Austin, 1962).

Austin's different classes of utterances or speech acts are based on his method of 'going through the dictionary' – i.e. to look up verbs in the dictionary and classify them according to their ordinary use (Austin, 1962; Chapman, 2005). This method is one of the reasons why Grice became dissatisfied with Austin – he found it unsystematic, random, and unfruitful as specific words are used with so many different senses that these senses do not necessarily have anything in common (Chapman, 2005).

Austin's emphasis on ordinary language and language use places him on the pragmatic side of language as opposed to Frege, for instance, who is on the semantic side of language. A crude distinction between semantics and pragmatics is that semantics is concerned with literal word- and sentence-meaning (i.e. context invariant meaning), reference⁴³, and truth conditions, whereas pragmatics is concerned with utterer's (speaker-) meaning (i.e. context dependent meaning), speech acts, indexicals, supposition, and in general all the aspects of language which are dependent on users and context (Austin, 1962; Davis, 2005; Fox, 1983; Frege, 1892; Grice, 1989; Levinson, 1983). The distinction is crude as there is no clear cut line between semantics and pragmatics. Where the line is drawn depends on the field (e.g. linguistics, philosophy of language, second language acquisition, etc.). However, the simplest and most common distinction between semantics and pragmatics is the distinction between word- and sentence-meaning (semantics) and speaker-meaning (pragmatics). In other words

"Pragmatics deals with *utterances*, by which we will mean specific events, the intentional acts of speakers at times and places, typically involving language. Logic and semantics traditionally deal with properties of *types* of expressions, and not with properties that differ from token to token, or use to use, or, as we shall say, from utterance to utterance, and vary with the particular properties that differentiate them." (Korta & Perry, 2006, section 1).

Austin and Grice are both influenced by Frege and his theory of meaning in *On Sense and Reference* (1892). Hence, Austin's definition of a locutionary act as an utterance "with a certain 'meaning' in the favourite philosophical sense of that word, i.e. with a certain sense and with a certain reference" (Austin, 1962, p. 94) is based on Frege's notions of 'sense' and 'reference'.

⁴³ It is possible to distinguish between semantic reference and speaker reference, i.e. to distinguish between what words and sentences in themselves refer to and what people refer to when they use those words and sentences. When this distinction is made 'semantic reference' belongs to semantics, whereas 'speaker reference' belongs to pragmatics.

In *On Sense and Reference* (1892) Frege explains why there is a difference between the sentences

(1) "The evening star is the evening star"

and

(2) "The evening star is the morning star"

even though both the 'Evening star' and the 'Morning star' refer to the planet Venus. In (1) the logical form is $a = a$, which is a tautological statement that does not tell anything. In (2), however, the logical form is $a = b$, which does provide new insights especially if the hearer of the sentence does not know or believe that a and b (the morning star and the evening star) are the same planet, i.e. Venus. Thus, "It is natural, now, to think of there being connected with a sign (name, combination of words, letter), besides that to which the sign refers, which may be called the reference of the sign, also what I should like to call the *sense* of the sign, wherein the mode of presentation is contained." (Frege, 1892, p. 143). That is, a given sign (e.g. 'Evening star') refers to some object as its reference (i.e. Venus) through a specific mode of presentation, its sense (i.e. 'Evening star' instead of 'Morning star'). Even though 'Evening star' and 'Morning star' designate the same reference (i.e. Venus) they express different senses among which humans can discriminate (Frege, 1892). Therefore, a human agent can discriminate between (1) and (2) in the sense that he can believe (1) without believing (2). If a given sign or proper name does not designate any real object, as for instance fictitious names (e.g. Zeus), then the proper name does not have a reference, but it still have a sense, which humans can grasp and understand (Frege, 1892). The description of sense and reference as 'the mode of presentation' and 'the object referred to' only hold for signs or proper names. When declarative sentences are considered the sense and reference are something else. In contrast to proper names, declarative sentences contain thoughts and they do not necessarily refer to only one object. The thought is the "objective content [of the performance of thinking], which is capable of being the common property of several thinkers." (Frege, 1892, p. 146) and can be understood as a proposition (McGrath, 2005). The thought contained in a given sentence, is the sense of that sentence. If the proper name Plato in the sentence

(3) "Plato wrote Theaetetos"

is replaced by another proper name with the same reference, such as

(4) "Aristotle's teacher wrote Theaetetos"

then the reference of the whole sentence (3) is unchanged in (4). However, the thought contained in (4) changed with the change of the proper name and is different from the

thought in (3), i.e. they express two different senses – they designate the reference in two different ways. The reference of a sentence is its truth-value (cf. section 4.1.1), i.e. that it is either true or false. Thus, “[e]very declarative sentence concerned with the reference of its words is therefore to be regarded as a proper name, and its reference, if it has one, is either the True or the False [designated as objects].” (Frege, 1892, p. 147). If a sentence contains a proper name which has no reference (e.g. Zeus) then the sentence as a whole has no reference either. However, the sentence

(5) “Zeus is a Greek god”

still contains a thought (i.e. expresses a sense) and is a perfectly understandable sentence although it has no reference (i.e. has no truth-value – is neither true, nor false). Frege’s own example is as follows: “The sentence ‘Odysseus was set ashore at Ithaca while sound asleep’ obviously has a sense. But since it is doubtful whether the name ‘Odysseus,’ occurring therein, has reference, it is also doubtful whether the whole sentence has one.” (Frege, 1892, p. 146). As Frege considers the truth-values True and False to be actual objects, a sentence which includes a fictitious name such as ‘Odysseus’ (or ‘Zeus’) cannot have a reference (not even within that work of fiction where it is presented). That fictitious (or non-existent) proper names do not have references in Frege’s theory, but only senses is the basis for Russell’s critique of the theory:

“Russell observes in his 1905 article ‘On denoting’ that ‘a phrase may be denoting, and yet not denote anything; for example, “the present king of France”’. For Frege such examples posed no problem. It was perfectly possible to say that such an expression had a sense but no reference (...). If ‘the king of France’ simply has no reference, then any sentence of which it is a subject, such as Russell’s famous example ‘the king of France is bald’, must also fail to refer to the world; in other words it must fail to be either true or false. For Russell this result is intolerable because it dispenses with classical, two-valued logic, in which every proposition must be capable of being judged either ‘true’ or ‘false’.” (Chapman, 2005, p. 35).

For Frege (1892), it is words and sentences themselves which express senses and designates references, not speakers. A reference (or truth-value) of a proper name (or sentence) is an object in the world and is therefore not influenced by what a speaker might think he refers to in uttering specific words or sentences. The meaning (i.e. sense) of a sentence is determined by the meaning (i.e. sense) of the words, signs, proper names within that sentence. It is the literal word- and sentence-meaning which Frege considers – i.e. it is purely semantics. Austin’s speech act theory, on the other hand, is purely pragmatics as it is solely language use, which determines meaning. Speech acts are dependent on speakers, hearers, and context –

even on the locutionary level. What an utterance means is determined by what the speaker means – i.e. which speech acts he performs of, in, and by saying what he says.

Although Grice is considered an ordinary language philosopher who thereby belongs to the realm of pragmatics there is more to his theories than descriptions of communicative practice:

“It is, in my view, an important part, though by no means the whole, of the philosopher’s task to analyze, describe, or characterize (in as general terms as possible) the ordinary use or uses of certain expressions or classes of expressions.” (Grice, 1958, p. 172).

Like Austin, Grice considers ordinary language and language use pivotal to philosophy, however, he also shares common grounds with Frege, in his search for, and description of, general principles and theories. Hence,

“[m]uch of his [Grice’s] philosophy of language, particularly his theory of conversational implicature, has for a number of decades played a central role in debates about the relationship between semantics and pragmatics, or meaning as a formal linguistic property and meaning as a process taking place in contexts and involving speakers and hearers.” (Chapman, 2005, p. 1).

and the distinction between sentence-meaning and utterer’s meaning (or speaker-meaning) and the insight that speakers can mean more with their utterances than the mere sentence-meaning “paved the way for the emergence of pragmatics as a separate discipline.” (Chapman, 2005, p. 188).

2.1.2 Grice’s position within philosophy of language

Grice’s approach to language is grounded in the belief that the way humans use language is rational, wherefore, there must be some principles or “rules”, which can explain why human linguistic behavior is indeed rational (Grice, 1967). According to Grice “it is just a well-recognized empirical fact that people do behave in these ways [i.e. cooperate in discourse]” (Grice, 1967, p. 29), however, Grice is “enough of a rationalist to want to find a basis that underlies these facts, undeniable though they may be” (Grice, 1967, p. 29). Grice’s description of himself as ‘enough of a rationalist’ is a reference to Kant and his categories for a priori understanding, which underlies Grice’s quest to go beyond the description of actual conversational practice in order to show why that practice is rational and reasonable (cf. sections 2.3 and 2.4). The combination of linguistic behavior and rational principles is the reason why Grice is neither a formalist nor an informalist – he acknowledges both ordinary language and technical language as relevant for philosophy:

“[Grice] displayed a sophisticated respect for common sense as a starting point in philosophical inquiry. It is sophisticated in that he did not espouse the straightforward adoption of common sense attitudes and terminology into philosophy. Rather, he urged that the philosopher remain aware of how the themes of philosophical inquiry are usually understood by non-philosophers. In a large part, this meant paying serious attention to the language in which particular issues were ordinarily discussed, or to what Grice once described as ‘our carefree chatter’.” (Chapman, 2005, p. 4).

In this formulation Grice’s view on the use of common sense in philosophy is an echo of Carnap’s (1950) method of explication (cf. section 1.4) where the explicandum (the term in its common sense) must undergo an explication (a transformation) into an explicatum, i.e. a fruitful concept for the philosophical or scientific domain where it is needed. In Carnap’s method the common sense or ordinary use of a term is thus the starting point for philosophical or scientific progress, with emphasis on ‘starting point’ as the explicandum, although necessary, is not sufficient in itself – the explicatum is needed as well.

2.2 The Gricean program – meaning as intention

Grice’s philosophy of language, which draws on formalism in terms of general principles as well as informalism in terms of ordinary language use, laid the foundation for his realization of the need to distinguish between word- and sentence-meaning and utterer’s meaning. The foundation of this distinction is laid in *Meaning* (1957), one of Grice’s earliest publications,⁴⁴ and is developed throughout his work with its most explicit formulation in *Logic and Conversation* (1967).

In the essay *Meaning* (1957) Grice develops a distinction between what he calls natural and nonnatural meaning. Natural meaning is defined by the formula “*x meant that p* and *x means that p* entail *p*” (Grice, 1957, p. 213). One of Grice’s own examples of natural meaning is

(1) “Those spots mean (meant) measles.”⁴⁵

where the measles are entailed by the occurrence of those exact spots in such a way that the spots cannot be present without measles being present as well. Thus, it is not possible to say

⁴⁴ *Meaning* is Grice’s second official publication. It was published in *The Philosophical Review*, 66, in July 1957. In *Studies in the Way of Words* (1989) Grice has added the year 1948 to *Meaning*, which indicates that it was written nine years prior to its publication.

⁴⁵ Grice, 1957, p. 213.

(2) "Those spots meant measles, but he hadn't got the measles"⁴⁶

neither is it possible to reformulate the sentence "Those spots meant measles" into

(3) "Those spot meant 'measles'"⁴⁷

where the inverted commas indicate that 'measles' is not to be understood in its ordinary sense or is to be understood as the term 'measles' rather than actual measles. Instead, it is possible to reformulate instances of natural meaning by use of "The fact that...", e.g.

(4) "The fact that he had those spots meant that he had measles."⁴⁸

In other words, in cases of natural meaning x and p are glued together and cannot be separated. Natural meaning is contrasted with nonnatural meaning which has the formula " x means that p and x meant that p do not entail p ." (Grice, 1957, p. 214). Grice's example of nonnatural meaning is

(5) "Those three rings on the bell (of the bus) mean that the bus is full."⁴⁹

In cases like this it is possible to add

(6) But in fact the bus is not full, it was a mistake.

And it is also possible to reformulate the statement with use of inverted commas such that it reads

(7) "Those three rings on the bell mean 'the bus is full'."⁵⁰

In cases of nonnatural meaning the x and the p are not glued together in the same way as with natural meaning. Therefore the sentence

(8) "The fact that the bell has been rung three times means that the bus is full."⁵¹

is not a restatement of (5) the same way as (4) is a restatement of (1), in natural meaning, as the two sentences (8) and (5) do not have the same meaning. Grice emphasizes that both (5) and (8) might be true but they differ in meaning because nonnatural meaning leaves room for

⁴⁶ Grice, 1957, p. 213.

⁴⁷ Grice, 1957, p. 213.

⁴⁸ Grice, 1957, p. 214.

⁴⁹ Grice, 1957, p. 214.

⁵⁰ Grice, 1957, p. 214.

⁵¹ Grice, 1957, p. 214.

agents who can make mistakes as is exemplified in (6). In nonnatural meaning there is room for agents to mean something by what they do and what they say. What these agents mean by their actions and utterances might be right or true or it might be wrong or false. Thus, the three rings on the bell on the bus can rightly mean that the bus is full on occasions where the chauffeur misinterprets the bus as being full and therefore rings the bell three times even though the bus is not full. The distinction between the natural sense and the nonnatural sense of meaning is further established when Grice describes which sentence-patterns fall under which sense of ‘meaning’:

“I propose, for convenience, also to include under the head of natural senses of “mean” such senses of “mean” as may be exemplified in sentences of the pattern “A means (meant) *to do* so-and-so (by x),” where A is a human agent. By contrast, as the previous examples show, I include under the head of nonnatural senses of “mean” any senses of “mean” found in sentences of the patterns “A means (meant) something by x” or “A means (meant) by x that...” (...).” (Grice, 1957, pp. 214-215).

That much said, Grice leaves natural meaning behind and concentrates on the question “What is meaning_{NN}?” (where ‘NN’ is an abbreviation for ‘nonnatural’) with the note that “[a]sking this question will not of course prohibit us from trying to give an explanation of “meaning_{NN}” in terms of one or another natural sense of “mean”.” (Grice, 1957, p. 215).

One of Grice’s key insights is that meaning is determined by how humans use language and this use is determined by the intentions which humans have when they communicate. Thus, for A (a human agent) to mean something (p) by x (an utterance), A must have three interdependent intentions. A must intend that

- (i) the hearer will be convinced that p,
 - (ii) that the hearer recognizes A’s intention in (i), and
 - (iii) that it is because of the recognition in (ii) that the hearer is convinced that p
- (Grice, 1957).

All three intentions are necessary for nonnatural meaning to be present. This is not to say that a human cannot utter⁵² something without the presence of these three intentions; he can, but then it is not an instance of nonnatural meaning.

In order to further explain how nonnatural meaning is determined by the three intentions of the utterer (or speaker) Grice introduces three generalizations:

⁵² Grice uses ‘utterance’ as a neutral word which can apply to verbal as well as non-verbal acts.

“(1) “A meant_{NN} something by x” is (roughly) equivalent to “A intended the utterance of x to produce some effect in an audience by means of the recognition of this intention” (...) to ask what A meant is to ask for a specification of the intended effect (...). (2) “x meant something” is (roughly) equivalent to “Somebody meant_{NN} something by x.” (...) I feel inclined to say that (as regard traffic lights) the change to red meant_{NN} that the traffic was to stop; but it would be very unnatural to say, “Somebody (e.g. the Corporation) meant_{NN} by the red-light change that the traffic was to stop.” Nevertheless, there seems to be *some* sort of reference to somebody’s intentions. (3) “x means_{NN} (timeless) that so-and-so” might as a first shot be equated with some statement or disjunction of statements about what “people” (vague) intend (...) to effect by x.” (Grice, 1957, p. 220).

The first generalization is a comprised description of the three intentions with the further elaboration that to ask for meaning is actually to ask for the intention behind what is said. In the second and the third generalizations it is not an agent who means_{NN} something but an object, a word, an event, etc. (i.e. an x) which acquires its meaning_{NN} through the use and intentions of utterers (i.e. agents). In this way the second and third generalizations are connected to conventions as the collective intentions of multiple agents. The reference to ‘somebody’s intentions,’ in the case with the traffic lights, is thus a reference to the convention that red light in the traffic means_{NN} stop. It is a convention, and an instance of nonnatural meaning, because it could have been different. The convention that red light means_{NN} stop is based on the three intentions that

- (i’) the agent is convinced to stop (by the sight of the red light),
- (ii’) that the agent recognizes the intention in (i) (i.e. to convince him to stop by the sight of the red light), and
- (iii’) that it is because of the recognition in (ii) (i.e. the recognition of the intention to convince him to stop) that the agent stops. In other words the agent recognizes and understands the convention behind (i.e. the meaning_{NN} of) the red light.

The emphasis on utterers’ intentions as that which explains meaning is one of Grice’s main contributions to philosophy of language. In his article *Theories of Meaning* (2010) Speaks refers to this contribution as the ‘Gricean program’:

“Paul Grice developed an analysis of meaning which can be thought of as the conjunction of two claims: (1) facts about what expressions mean are to be explained, or analyzed, in terms of facts about what speakers mean by utterances of them, and (2) facts about what speakers mean by their utterances can be explained in terms of

their intentions. These two theses comprise the ‘Gricean program’ for reducing meaning to the contents of the intentions of speakers.” (Speaks, 2010, section 3.1.1)

In this way, *Meaning* (1957) laid the foundation for the development of the distinction between word-meaning, sentence-meaning, and speaker-meaning (with speaker-meaning as the more fundamental (Speaks, 2010)), which are at the core of another of Grice’s famous essays *Logic and Conversation* (1967).

2.3 Logic and Conversation

In the definitions of lying, misleading, and deceiving (cf. section 1.1) especially one notion proved crucial in the distinction between lying, on the one hand, and misleading and deceiving, on the other hand, namely ‘implicature’ (more precisely false implicature). This notion was put forth by Grice in the essay *Logic and Conversation* (1967) along with his Cooperative Principle and its maxims. According to the literature on lying, misleading, and deceiving (see for example Adler, 1997; Fallis, 2010; Mahon, 2008; Stokke, 2013; and Webber, 2013), it is through the use of implicatures that people are able to verbally mislead or deceive one another without merely lying. ‘Misleading’ and ‘deceiving’ – and thereby ‘implicature’ – are essential in order to explore and understand the notions of misinformation and disinformation (cf. sections 1.2 and 1.5) as implicatures are generated when maxims are exploited or flouted while the Cooperative Principle is observed.

Logic and Conversation (1967) is perhaps the most influential and most famous piece of Grice’s work. A reference to *Logic and Conversation* can point to two different entities. It can refer to the collective name of Grice’s William James lectures from 1967 as well as the title of the second lecture published as an individual essay in 1975. After the delivery of the William James lectures at Harvard University the lectures were circulated in mimeograph and were published individually between 1968 and 1989⁵³. The published essays and the unpublished lectures from The William James lectures were later collected and partly revised to “form a centerpiece” in Grice’s posthumously published collection of essays *Studies in the Way of*

⁵³ The first lecture *Prolegomena* was first published in *Studies in the Way of Words* (SWW) in 1989. The second lecture *Logic and Conversation* was published in 1975 in Cole, P. and Morgan, J.L. (eds.). *Syntax and semantics 3: Speech Acts*, as well as in Davidson, D. and Harman, G. (eds.). *The Logic of Grammar*, also in 1975. The third lecture *Further Notes on Logic and Conversation* was published in 1978 in Cole, P. (ed.). *Syntax and semantics 9: Pragmatics*. The fourth lecture *Indicative Conditionals* was first published in SWW. The fifth lecture *Utterer’s Meaning and Intentions* was published in 1969 in *Philosophical Review* 78. The sixth lecture *Utterer’s Meaning, Sentence-Meaning, and Word-Meaning* was published in 1968 in *Foundations of Language* 4. And the seventh lecture *Some Models for Implicature* was first published in SWW (Chapman, 2005, pp. 235-236, Grice, 1989).

Words (1989) (Grice, 1989, p. v; Chapman, 2005). In *Studies in the Way of Words* Grice has dated all the essays, which comprise the William James lectures to 1967, with revision in 1987.

2.3.1 The Cooperative Principle and its maxims

In the essay *Logic and Conversation* (1967) Grice presents what is known today as the conversation theory. The theory begins “with a characterization of the notion of “implicature”” (Grice, 1967, p. 24) which leads to the formulation of the Cooperative Principle (CP) and its four maxims. The Cooperative Principle and the maxims are then used to investigate why and how implicatures arise in conversation.

Grice observes that

“[o]ur talk exchanges do not normally consist of a succession of disconnected remarks, and would not be rational if they did. They are characteristically, to some degree at least, cooperative efforts; and each participant recognizes in them, to some extent, a common purpose or set of purposes, or at least a mutually accepted direction.” (Grice, 1967, p. 26)

From these observations Grice formulates a general principle which governs conversational enterprise as he has observed it. This general principle is the Cooperative Principle which says:

“Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.” (Grice, 1967, p. 26).

It is a very simple principle which spells out the rules which underlies a successful conversation. Even though the principle is stated in imperative it describes the expectations the participants in a conversation have of each other as cooperative in their shared endeavor. Under the Cooperative Principle falls several different maxims divided into four different categories: Quantity, Quality, Relation, and Manner⁵⁴. These maxims further specify the otherwise tacit conversational rules which people normally adhere to.

⁵⁴ This is a direct reference to Kant: “Echoing Kant, I call these categories Quantity, Quality, Relation, and Manner.” (Grice, 1967, p. 26).

Quantity (how much information is to be provided):

1. Make your contribution as informative as is required (for the current purposes of the exchange).
2. Do not make your contribution more informative than is required.

Quality (of the information provided):

Supermaxim – Try to make your contribution one that is true

1. Do not say what you believe to be false.
2. Do not say that for which you lack adequate evidence.

Relation:

1. Be relevant

Manner (the form of the utterance):

Supermaxim – Be perspicuous

1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief (avoid unnecessary prolixity):
4. Be orderly.

Fig. 2.1 – Maxims (Grice, 1967, pp.26-27).

The maxims within the categories Quantity, Quality, and Relation are related to the content of utterances – i.e. *what* is said – whereas the maxims within the category Manner relate “to *how* what is said is to be said” (Grice, 1967, p. 27). Although the maxims are stated in imperative they shall be read as a ‘formal’ way of stating the “rules” which already govern conversation:

“... what I regard as a fundamental question about the Cooperative Principle and its attendant maxims, namely, what the basis is for the assumption which we seem to make, and on which (I hope) it will appear that a great range of implicatures depends, that talkers will in general (*ceteris paribus* and in the absence of indications to the contrary) proceed in the manner that these principles prescribe. A dull but, no doubt at a certain level, adequate answer is that it is just a well-recognized empirical fact that people do behave in these ways;(…) I am, however, enough of a rationalist to want to find a basis that underlies these facts, undeniable though they may be; I would like to be able to think of the standard type of conversational practice not

merely as something that all or most do *in fact* follow but as something that it is *reasonable* for us to follow, that we *should not* abandon.” (Grice, 1967, p. 29).

Even though, the Cooperative Principle and its maxims are Grice’s formulation of the underlying “rules” which people adhere to in conversation these maxims can be – and frequently are – violated and exploited in different ways. Grice describes four different ways in which a speaker may fail to fulfill a maxim (Grice, 1967, p. 30):

1. He may simply and quietly *violate* a maxim in which case he will be liable to mislead as the violation is not obvious.
2. He may *opt out* from the cooperation by not be willing to observe the Cooperative Principle or the maxims.
3. There might be a *clash* such that it is not possible to fulfill one maxim without the violation of another maxim.
4. The speaker may *flout* a maxim by “blatantly fail to fulfill it.”.

An example of the *violation* of a maxim is the telling of a lie, as this is a violation of the maxims of Quality, both the supermaxim (“Try to make your contribution one that is true”) and the first maxim (“Do not say what you believe to be false”). An example of *opting out* is when a person openly refuses to continue or take part in a conversation by simply not answering or by saying something along the lines of “I will not answer that” or “I cannot say any more.” A clash happens, for instance, when a speaker might not be able “to fulfill the first maxim of Quantity (Be as informative as required) without violating the second maxim of Quality (Have adequate evidence for what you say).” (Grice, 1967, p. 30). Perhaps the most interesting way in which a speaker can fail to fulfill a maxim is by *flouting* it. The flouting of a maxim is interesting because in this case it is actually possible for the speaker to fulfill the maxim in question, however, he chooses not to do so. The choice not to fulfill the maxim leaves the hearer with a puzzle and a question: “Why did the speaker choose to flout the maxim instead of fulfilling it?”

If a speaker flouts a maxim, that is he fails to fulfill the maxim even though he is able to fulfill it (i.e. he is not faced with a clash, he is not opting out, he is not trying to mislead) and he, at the same time, observes the Cooperative Principle then a conversational implicature is generated and the maxim is *exploited*.⁵⁵

⁵⁵ In Grice’s (1967) descriptions of the four different ways in which a speaker can fail to fulfil a maxim, on p. 30, it may appear as if it is only the flouting (and thereby the exploitation) of a maxim that generates an implicature.

2.3.2 Implicature

The notion of ‘implicature’ is a pragmatic notion in the sense that it denotes what a speaker, or an utterer in Grice’s vocabulary, means or implicates by an utterance beyond what he literally says. To make an implicature is to implicate something by what is said without literally saying it – that is to mean something more or something different than what is literally said – and it is dependent on the context of the utterance: who says it?; who or what is it about?; when is it said?; in what connection is it said?; etc. To ‘implicate’ is the act of including more or something else than what is literally said, and ‘implicatum’ is that which is implicated⁵⁶. Implicatures can be of two different kinds; either what is implicated can lie within the utterance itself and be determined by the conventional meaning of the words, or what is implicated can lie beyond the utterance itself. In the first case the implicature is called conventional – it can be derived from the utterance itself, without knowledge of the context, solely on the basis of the conventional word-meaning:

“If I say (smugly), *He is an Englishman; he is, therefore, brave*, I have certainly committed myself, by virtue of the meaning of my words, to its being the case that his being brave is a consequence of (follows from) his being an Englishman. But while I have said that he is an Englishman, and said that he is brave, I do not want to say that I have *said* (in the favored sense) that it follows from his being an Englishman that he is brave, though I have certainly indicated, and so implicated, that this is so.” (Grice, 1967, p. 25).⁵⁷

The implicature is conventional because what is implicated (i.e. what is meant) is fixed by the conventional meaning of the words and not solely by what the utterer means by what he says, as in the case of conversational implicatures. In conversational implicatures, what is implicated lies beyond the utterance itself. It is determined by what the utterer means and can only be understood when the context is taken into account:

“A is writing a testimonial about a pupil who is a candidate for a philosophy job, and his letter reads as follows: “Dear Sir, Mr. X’s command of English is excellent, and his

However, in the examples of conversational implicatures in the following pages all kinds of violations of the Cooperative Principle and its maxims can generate implicatures.

⁵⁶ “I wish to introduce, as terms of art, the verb *implicate* and the related nouns *implicature* (cf. *implying*) and *implicatum* (cf. *what is implied*).” (Grice, 1967, p. 24).

⁵⁷ Grice does not say more about conventional implicature than this and it is only the conversational implicatures which are “being essentially connected with certain general features of discourse [i.e. the Cooperative Principle and its maxims].” (Grice, 1967, p. 26).

attendance at tutorials has been regular. Yours, etc.” (Gloss: A cannot be opting out, since if he wished to be uncooperative, why write at all? He cannot be unable, through ignorance, to say more, since the man is his pupil; moreover, he knows that more information than this is wanted. He must, therefore, be wishing to impart information that he is reluctant to write down. This supposition is tenable only if he thinks Mr. X is no good at philosophy. This, then, is what he is implicating.)” (Grice, 1967, p. 33).

The conversational implicature that Mr. X is no good at philosophy is not generated by the utterance from the letter in itself. Taken literally the utterance is a praise of Mr. X’s skills in English as well as his capability of showing up when he has to. The implicature is generated by that specific utterance in the specific context of a testimonial for a job application because the utterance does not provide what is expected from such a testimonial, namely a description or praise of Mr. X’s skills as a philosopher. Hence, A flouts the maxim of Relation (“Be relevant”) as well as the first maxim of Quantity (“Make your contribution as informative as is required.”) as the utterance in the letter is not strictly relevant for the purpose and contains less information than required. The writer means something else and something more than what he literally writes.

The introduction of ‘implicature’ is thus an explicit introduction of Grice’s famous distinction between sentence-meaning (i.e. what is literally said) and utterer’s meaning (i.e. what is implicated) – the distinction which was alluded to in *Meaning* (1957) (cf. section 2.2).

Implicatures are generated when the overall Cooperative Principle is observed but one or more of the maxims are exploited or violated (although there are “[e]xamples in which no maxim is violated, or at least in which it is not clear that any maxim is violated” (Grice, 1967, p. 32)). Exactly how this violation or exploitation has to come about is not exactly clear in Grice’s writing (cf. note 55). In the presentation of the four different ways in which maxims can be violated or exploited it appears as if only the flouting of a maxim (while observing the Cooperative Principle) causes the generation of a conversational implicature. However, in the various examples of conversational implicatures, these can be generated without any clear violation of a maxim; by violation due to a clash between maxims; or by “*exploitation, that is, a procedure by which a maxim is flouted*” (Grice, 1967, p. 33). Thus, any kind of violation of a maxim generates a conversational implicature as long as the Cooperative Principle is observed – or assumed to be observed – by the utterer and as long as the implicature can be – and have to be – worked out: “The presence of a conversational implicature must be capable of being worked out; for even if it can in fact be intuitively grasped, unless the intuition is replaceable by an argument, the implicature (if present at all) will not count as a conversational implicature; it will be a conventional implicature.” (Grice, 1967, p. 31).

‘Implicature’ is a mechanism of voluntary exploitation or violation of a maxim as a way to say something more or something else, than what is literally said, while still being cooperative. People generate implicatures for all sorts of reasons and sometimes the generation of an implicature is a way to uphold the cooperation – for example by being polite instead of completely honest, as when someone asks “Did you like the book I recommended?” and one answers “It has a nice cover” with the implicatum: “No, I didn’t like the book, but in order to be polite I say the best thing about it, that I possibly can, namely, that it has a nice cover.”

2.3.3 False implicature, to work it out, and the possibility of mistakes

It is this feature of implicature (i.e. the implication of something else than what is literally said) which the definitions of ‘misleading’ and ‘deceiving’ as false conversational implicatures are based upon (cf. section 1.1). A false implicature is an implicature where what is implicated, i.e. the implicatum, is false while what is literally said is true. “Since the truth of a conversational implicatum is not required by the truth of what is said (what is said may be true – what is implicated may be false), the implicature is not carried by what is said, but only by the saying of what is said, or by “putting it that way.”” (Grice, 1967, p. 39). The saying of a given expression always occurs within a context of time and space and often within a conversation as well, which means that it occurs in connection to, and as a consequence of, other expressions. This is the reason why it is not what is said which carries the conversational implicature but instead the act of saying what is said, in that specific context, which carries the implicature – it is generated by the utterer’s communicative intentions. That conversational implicatures are carried or generated ‘by saying what is said’ is one of the reasons why they have to be worked out by a hearer (the hearer has to infer what is implicated; Bach, 2006). In order

“[t]o work out that a particular conversational implicature is present, the hearer will rely on the following data: (1) the conventional meaning of the words used, together with the identity of any references that may be involved; (2) the Cooperative Principle and its maxims; (3) the context, linguistic or otherwise, of the utterance; (4) other items of background knowledge; and (5) the fact (or supposed fact) that all relevant items falling under the previous headings are available to both participants and both participants know or assume this to be the case.” (Grice, 1967, p. 31).

Grice repeatedly emphasizes that the hearer supposes or assumes that the speaker observes the conversational maxims, or at least the Cooperative Principle, and that the speaker is presumed to do so in order to implicate something by his utterance (Grice, 1967). Despite this emphasis on supposition and assumption, Grice neither considers how – or if – it can be known that a speaker observes the Cooperative Principle and its maxims, nor what happens if a hearer incorrectly assumes that a speaker observes the Cooperative Principle and its

maxims. Therefore, it is not clear whether it is possible for a speaker to accidentally generate a conversational implicature simply on the grounds that a hearer works out an implicatum from the speaker's utterance on the false/incorrect assumption that the speaker observes the Cooperative Principle and thus flouts a maxim on purpose (and not – which is the actual case in this example – by accident). However, the addition of 'supposed fact' in the fifth kind of data which the hearer relies on in the working-out process – and the fact that an implicatum is never said or stated – seems to leave open the possibility for mistakes.

The mistake and thereby the accidental generation of an implicature may be caused by a performance error on the part of the utterer. Competent speakers of English and competent Gricean contributors may once in a while – due to stress, laziness, or fatigue – make performance errors that produce unintended implicatures 'by accident'. The mistake can also be on the part of the hearer who might work out the 'wrong' implicature (or work out an implicature where none was intended) and arrive at an implicatum which the utterer had not in mind. The working-out process also enables an utterer – who did in fact make a false conversational implicature – to deny that he made a false implicature. Such an utterer can employ one of two strategies in order to deny his attempt to mislead: either he can claim that the hearer draws false conclusions (i.e. works out a 'wrong' implicatum); or he can pretend to be ignorant about what his utterance could implicate (i.e. that an implicature is present). In both cases the utterer can simply say "That's not what I meant" and, if he wants to, he can continue with a literal statement of the intended implicatum "I meant that..." or with a restatement of what he literally said which does not generate an implicature.

According to Bach (2006) there can be no such thing as accidental implicatures as "what a speaker implicates is a matter of his communicative intention in uttering the sentence. That's why implicature is pragmatic in character, hence why in different situations one can utter a given unambiguous sentence and implicate different things." (Bach, 2006, p. 22). The cases of performance errors and hearers who work out (i.e. infer) implicatures where none were intended are cases of putative implicatures in Bach (2006) not actual implicatures:

"people sometimes confuse *infer* with *implicate* and *inference* with *implicature*. Why is this difference important? One obvious reason is that the audience can take the speaker to be implicating something when in fact he isn't. A putative implicature need

not be an actual one. Equally obviously, a speaker can implicate something even if the audience doesn't make the intended inference." (Bach, 2006, p. 23).⁵⁸

This means that only intended implicatures are actual implicatures – unintended 'implicatures' are not actually implicatures but something else. However, Grice does not explicitly address the possibility of mistakes in *Logic and Conversation* (1967) as he solely considers intended violations or exploitation of the maxims. In the literature on lying, misleading, and deceiving (cf. section 1.1) the generation of false implicatures is always intentional with the exception of Mahon (2008) where 'misleading' is un-intentional as opposed to 'deception', which is intentional. In the case of un-intentional misleading the utterer might actually be ignorant about what his utterance could falsely be taken to implicate (i.e. he violates a maxim unknowingly) but that does not make the utterance less misleading it just was not intended to mislead.

2.4 Influence and critique

Logic and Conversation (1967) – and *Meaning* (1957) – has had, and still have, a great influence in philosophy of language, in linguistics, in philosophy of information, and in the philosophical literature on lying/misleading/deceiving (cf. section 1.1), and "[m]ost current pragmatic theorists are neo-Griceans in that they adopt at least some version of his main three contributions" (Korta & Perry, 2006, section 3).

The three main contributions are (Korta & Perry, 2006):

- the distinction between what an utterer literally says and what the utterer implicates (*Meaning* and *Logic and Conversation*);
- that some rules or principles govern or guide communication (*Logic and Conversation*); and
- the intentions present in connection to nonnatural meaning which has to be recognized by the addressee (*Meaning*).

The influence of Grice's work – especially his insight on the differences between literal meaning and utterer's meaning – has paved the way for 'Gricean Pragmatics':

"[a] term [which] has been used as a label for a disparate body of work. It includes relatively uncritical applications of Grice's ideas to a wide range of different genres

⁵⁸ "Inference (...) is the activity performed by a reader or an interpreter in drawing conclusions that are not explicit in what is said." (Bach, 2006, p. 23).

(...) These different types of work do not all agree on the efficacy of the Cooperative Principle, on the exact characterization of the maxims, or even on the nature of conversational implicature itself. What they have in common, qualifying them for the title 'Gricean', is an impulse to distinguish between literal meaning and speaker meaning, and to identify certain general principles that mediate between the two. The latter is perhaps Grice's single greatest intellectual legacy." (Chapman, 2005, p. 185).

Despite this influence and intellectual legacy Grice's work in *Logic and Conversation* (1967) has also been severely criticized for being idealized, normative, and prescriptive. "The theory of conversation [Cooperative Principle and maxims] has frequently been accused of presenting an unrealistic picture of human interaction, and even of attempting to lay down rules of appropriate conversational etiquette." (Chapman, 2005, p. 190). For instance, Fairclough developed his critical discourse analysis as a reaction to ordinary discourse analysis because he thought it idealized cooperative conversation, something for which he held Grice partly responsible (Chapman, 2005). Grice's Cooperative Principle and its maxims, however, are descriptive, but there is also a Kantian element as the "rules" describe the conditions under which communication is possible. This is what Grice is getting at, when he writes that he is "enough of a rationalist" (Grice, 1967, p. 29) to want to explain why it is reasonable to "follow the standard type of conversational practice" (Grice, 1967, p. 29; cf. section 2.3.1) as well as, name his maxims after Kant's categories. The conception that *Logic and Conversation* is a normative work rather than a descriptive work is most likely caused by the fact that the Cooperative Principle and its maxims are stated in imperative – a form which underlines Grice's overall purpose with the lecture/essay – i.e. to formalize peoples' actual practice in communication as well as to find principles for why this practice is rational. Grice does use idealizations but they "are of the scientific type, necessary to establish underlying principles of the subject matter, rather than of the type specifying that it always does, or should, conform to certain optimal standards." (Chapman, 2005, p. 192). That it is the scientific type of idealizations which are at play is established by the outset of *Logic and Conversation* which is the divide, and the arguments, between formalists and informalists. It is Grice's wish

"to maintain that the common assumption of the contestants [formalists and informalists] that the divergences do in fact exist is (broadly speaking) a common mistake, and that the mistake arises from inadequate attention to the nature and importance of the conditions governing conversation. I shall, therefore, inquire into the general conditions that, in one way or another, apply to conversation as such, irrespective of its subject matter." (Grice, 1967, p. 24).

The divergences mentioned refer to the assumed divergences in meaning between the logical connectives and quantifiers and their analogues/counterparts in natural language. Thus, the Cooperative Principle does not prescribe how a conversation ought to proceed and the Cooperative Principle and its maxims are not rules which people ought to adhere to when they engage in conversation. Rather, the Cooperative Principle describes how conversations normally proceed and why this procedure is rational. Furthermore, the Cooperative Principle and its maxims are descriptions of the underlying “rules” of this normal procedure, which enable successful conversation – i.e. a way of spelling out the tacit rules which people actually adhere to in conversation and communication. In the same way the notion of implicature (both conventional and conversational) is a description of what happens when maxims are violated or exploited and why this violation or exploitation is rational.⁵⁹

2.5 Gricean framework

As mentioned, Grice’s theories of meaning and cooperation – with emphasis on natural and nonnatural meaning, the Cooperative Principle, the maxims, and conversational implicature – will serve as the theoretical framework in this dissertation. That is, a framework for discussions of the asymmetries in terms of meaning, truth, and intentions in the definitions of information, misinformation, and disinformation within philosophy of information. Furthermore, Grice’s development of the distinction between word-/sentence-meaning and utterer’s meaning is crucial for discussions of information, misinformation, and disinformation in communication, either in online social networks or in ordinary conversation. Besides being a fruitful framework for analyses and discussions of information, misinformation, and disinformation in general, in some way or other, Grice and his theories from philosophy of language are already present within the theories of information, misinformation, and disinformation presented in Ch. 1. In Dretske’s theory of semantic information it is Grice’s notion of meaning and his distinction between natural and nonnatural meaning which is present (cf. Ch. 3). For Dretske, information is natural meaning and it is because natural meaning is actual states of affairs that information is true. Floridi uses Grice – and Dretske – as one of the arguments why semantic information is veridical (cf. Ch. 4). Fallis uses Grice’s notion of implicature to argue in favor of the possibility of true disinformation (cf. Ch. 5), an argument based on the lying/deceiving/misleading distinctions (cf. Ch. 1). Furthermore, Fallis uses Grice’s Cooperative Principle incl. maxims to analyze different kinds of disinformation, as well as deception in general, where these kinds are classified according

⁵⁹ Grice does not recommend that implicatures should be avoided.

to which norm⁶⁰ they violate. Last, Fox uses the Cooperative Principle as an explanation for why most people conceive information as true when the concept does not require truth (cf. Ch. 4). This presence of Grice's work within the theories of information, misinformation, and disinformation in itself requires a reading of *Meaning* (1957) and *Logic and Conversation* (1967) in order to understand and discuss these theories of information, misinformation, and disinformation. The reading is required because Grice's theories have been applied to a vast range of research topics (from studies of metaphor, the language of jokes, and teaching of literacy, to historical linguistics, clinical linguistics, and language impairment, etc.) with more or less success as the theories in some cases have been stretched beyond their capability (Chapman, 2005).

Moreover, the application of Grice's theory of conversation to Artificial Intelligence is of special interest given the scope of the detecting-projects and serves as a further justification for a Gricean framework:

"Another field in which the theory of conversation has been used enthusiastically is Artificial Intelligence, a striking application given Grice's personal animosity to computers. Computer programmers have realized that formal linguistic theories can go a long way towards specifying the rules that make up grammatical sentences, but have little to say about how those sentences can be used to form natural-sounding conversations. For instance, Ayse Pinar Saygina and Ilyas Ciceki are not impressed by the pragmatic skills of computers. They analyse the output of various programmes entered for the Loebner Prize, an annual competition to find the programme closest to passing the Turing Test, in which a computer is deemed to be intelligent if a human observer cannot distinguish its participation in a conversation from that of a human. Saygina and Ciceki argue that pragmatic acceptability, rather than the simple ability to generate natural language, is the biggest challenge to such programmes. Moreover, the programmes that are the most successful are those that adhere most closely to Grice's maxims, suggesting future applications for these, specifically that programmers 'will inevitably have to find ways of "making computers cooperate"'. Various programmers have in fact made some attempts in this direction. Robert Dale and Ehud Reiter consider an algorithm for natural-sounding referring expressions in computer-human interaction. They note that such an algorithm must produce

⁶⁰ In the lying/misleading/deceiving literature the maxims are called norms although Grice did not depict them as norms.

expressions that successfully pick out the referent without leading ‘the human hearer or reader to make false conversational implicatures’.” (Chapman, 2005, pp. 195-196).

The realization that the closer a computer program adheres to Grice’s maxims the more successful it is in connection to the Turing Test is of interest in connection to the detecting-projects and the motivation for a Gricean framework (the aspect of communication as a cooperative enterprise) – it strengthens Grice’s theory of conversation as a description of actual practice and why this practice is rational. Furthermore, the observation, that algorithms must not cause humans to infer false conversational implicatures⁶¹ (even though these might only be putative), is interesting in connection to Floridi’s information-theoretic and computer scientific approach to information almost devoid of agents. It suggests, that truth in itself is not enough for humans to obtain knowledge, it must be secured in some way that the information is not misleading.

⁶¹ If one follows Bach (2006) then the quote “without leading ‘the human hearer or reader to make false conversational implicatures’.” (Chapman, 2005, p. 196) falls short to the misconception that implicatures can be described as inferences. According to Bach implicatures can only be generated by speakers and are always intended. Hearers and readers can only infer actual implicatures if these are present (i.e. intended by the speaker) – if a hearer wrongly infer an implicature where none was intended, then it is not an actual implicature (Bach, 2006). However, the point in the quote is simply that algorithms must be constructed in a way such that they do not mislead humans into inferring something false – which does not conflict with Bach (2006).

Ch. 3 Meaning

3 Meaning

Meaning as a concept of philosophical interest in itself has its place in philosophy of language on the boarder to linguistics (cf. Ch. 2). The debates about meaning concern how signs, words, sounds, etc. acquire meaning and what this meaning consists in. Various theories of meaning have been offered and meaning has, for instance, been described as the thoughts and ideas in people's minds conveyed through language, or as a reference-relation between words/sentences and objects/states of affairs (Collin & Guldman, 2005). Furthermore, the debates concern what it is to know or understand the meaning of some name, word, expression, or sentence. Is it, for example, to have a mental image of the meaning of some word or sentence, e.g. to have a mental image of redness? Or is it to be able to determine the reference of a word or sentence? (Collin & Guldman, 2005). The semantics-pragmatics distinction (cf. section 2.1.1) is at play here in connection to whether meaning is use and to know how to use different words and expressions (pragmatics, Grice) or whether meaning is the truth conditions of sentences and to know the truth conditions for a given sentence (semantics, Frege).

The notion of meaning is essential within most of the disciplines that comprise the philosophy of information as questions about content⁶², communication, learning, truth and falsity, etc. are all dependent on meaning in some sense – either in a strictly semantic sense or in both a semantic and a pragmatic sense. Within these various disciplines it is not the notion of

⁶² Content, or semantic content, is often used as equivalent to meaning.

meaning itself which is discussed. Instead specific notions of meaning from philosophy of language are adopted and used for specific purposes within different theories of, for example, information. Alternatively it is discussed how meaning is related to or constitutive of information. With regard to the accounts of information the disputes, in terms of meaning, concern whether meaning is a constitutive part of information or whether information exists before, and/or independent of meaning. It is these disputes, and their counterparts for the notions of misinformation and disinformation, which are the scope of this chapter. Thus, the aim is not to discuss the notion of meaning itself. Rather, the aim is to discuss which role 'meaning' plays in the various theories of information, misinformation, and disinformation in order to analyze and discuss the connections between these three notions. As part of the discussions and analyses on the role of 'meaning' it will be identified which theory of meaning is present within the various accounts – i.e. what is meant by 'meaning', semantics or pragmatics?

Although a theoretical framework in terms of meaning has already been laid out as the overall framework for this thesis (cf. Ch. 2) a rough distinction between different kinds of theories of meaning is essential in order to grasp which implications it has for a theory to employ a specific notion of meaning.

3.1 Theories of meaning

In the article *Theories of Meaning* (2010) on Stanford Encyclopedia of Philosophy, Jeff Speaks characterizes the various theories of meaning along a two-fold axis into two different kinds of theories: The semantic theories vs. the foundational theories of meaning. According to Speaks "a *semantic theory* – is a specification of the meanings of the words and sentences of some symbol system. Semantic theories thus answer the question 'What is the meaning of this or that expression?'" (Speaks, 2010, section 1) whereas "a *foundational theory of meaning* – tries to explain what about some person or group gives the symbols of their language the meanings that they have." (Speaks, 2010, section 1). Furner (2004) roughly makes the same distinction when he asks "What is it to say (i) that something has meaning (i.e., is meaningful) and (ii) that something has the particular meaning *p*?" (Furner, 2004, p. 430) where (i) alludes to foundational theories of meaning, whereas (ii) alludes to semantic theories. Within these two kinds of theories, which tries to answer different questions about meaning, different positions can be held. For the semantic theories a division between propositional and non-propositional theories can be made and for the foundational theories a division between mentalist and non-mentalist theories can be made (Speaks, 2010).

Within the semantic theories the debates concern what it is which constitutes the meaning of a given word, sentence, or expression. Both propositional and non-propositional semantic theories are offered as an alternative or an opposition to the referential theory of meaning.

The referential theory, or the theory of reference, simply states that the meaning of a word or sentence is the reference of that word or sentence. However, the referential theory encounters trouble when envisaged with:

- empty names (which have no reference but seem meaningful nonetheless),
- propositional attitude reports or belief ascriptions (where the reference is the same but the meaning changes), and
- indexicals (words such as ‘I’, ‘here’, ‘now’, and the like where the reference is ambiguous and context-dependent).

(Collin & Guldman, 2005; Furner, 2004; and Speaks, 2010).

According to Speaks “propositional theories supplement this theory of reference with an extra layer – with a theory which assigns a content, as well as a reference, to each meaningful expression.” (Speaks, 2010, section 2.2). Within these theories meaning is conceived as entities and the question is what counts as such an entity.

Frege’s proposal in *On Sense and Reference* (1892) is that the meaning of a word or a sentence is its sense, i.e. the way in which the word or sentence designates its reference. The sense of a sentence is the thought (or proposition) it expresses (cf. section 2.1.1). According to Harnish (1994) “Frege is the single most important philosopher of language” (Harnish, 1994, p. xx) and “almost always a point of departure. Theories in the philosophy of language are constantly being measured against the Fregean standard.” (Harnish, 1994, p. xx). Russell’s theory of definite descriptions in *On Denoting* (1905) (cf. section 2.1.1) is no exception as it contains a critique of Frege’s theory of sense and reference (Chapman, 2005; Harnish, 1994; and Russell, 1905).

In Russell’s theory meaning is also tied to propositions “[t]his is the principle of the theory of denoting I wish to advocate: that denoting phrases never have any meaning in themselves, but that every proposition in whose verbal expression they occur has a meaning.” (Russell, 1905, p. 162).⁶³ In contrast to Frege, Russell will not accept that sentences that contain empty names (or definite descriptions, i.e. ‘the king of France’) do not have a truth-value. Therefore, Russell proposes that expressions which include definite descriptions “are denoting phrases

⁶³ Russell, in a footnote, writes that “In this theory [Frege’s theory], we shall say that the denoting phrase expresses a meaning; and we shall say both of the phrase and of the meaning that they *denote* a denotation. In the theory, which I advocate, there is no *meaning*, and only sometimes a *denotation*.” (Russell, 1905, p. 173, note 10). This is part of his critique of Frege’s theory and ‘*meaning*’ and ‘*denotation*’ do not refer to the concepts in general, but are Russell’s notation for Frege’s ‘*sense*’ and ‘*reference*’. Thus, there is no equivalent to Frege’s ‘*sense*’ in Russell’s theory.

which do not denote anything.” (Russell, 1905, p. 171). Instead, an expression such as ‘the king of France is bald’ is a compound of various propositions (one of which is a proposition where the denoting phrase ‘the king of France’ occurs) “relating to the existence, the uniqueness, and the characteristics, of the individual apparently identified. Russell’s rendition of this example can be paraphrased as: ‘there exists one entity which is the king of France, and that entity is unique, and that entity is bald.’” (Chapman, 2005, p. 35). As there is no actual king of France and as ‘the king of France’ is the primary occurrence in the sentence, the first proposition is false, wherefore, the whole expression or sentence is false (Russell, 1905). Russell’s theory of denoting is tied to truth as his motivation for the development of the theory is to maintain bivalent logic (Chapman, 2005). Thus, “out of any proposition we can make a denoting phrase, which denotes an entity if the proposition is true, but does not denote an entity if the proposition is false. (...). The whole realm of non-entities (...) are denoting phrases which do not denote anything.” (Russell, 1905, p. 171).

In non-propositional semantic theories, meaning is not dependent on propositions and is not an entity. Thus, Davidson’s truth-conditional theory of meaning, where the meaning of a sentence is its truth-conditions (based on Tarski’s theory of truth, cf. section 4.1.2) is an example of such a theory (Collin & Guldman, 2005; Speaks, 2010).

Within the “foundational theories of meaning, which are attempts to specify the facts in virtue of which expressions of natural languages come to have the semantic properties that they have” (Speaks, 2010, section 3) the debates concern whether meaning should be explained in terms of mental states of language users (the mentalist approach) or whether meaning is solely to be explained by use of expressions without any references to mental representations (the non-mentalist approach) (Speaks, 2010).

The mentalist approach is a reductionist approach, where linguistic meaning is reduced to the content of mental states. Different candidates for which mental states (or representations) are the determinant for linguistic meaning have been proposed, for example, beliefs and intentions (Speaks, 2010). Grice’s theory in *Meaning* (1957) is a prime example of a foundational mentalist theory as it explains meaning in terms of the intentions of utterers (cf. Ch. 2). Another example is “the view that the meanings of expressions are fixed by conventions which pair sentences with certain beliefs.” (Speaks, 2010, section 3.1).

In the non-mentalist approaches language use is also determinant of meaning, however, it is not the use in terms of which beliefs or intentions the users might have with their utterances when they utter them. Rather, it is use in a historical sense – the use over time – which assigns a name, a word, an expression, etc. with a meaning (Speaks, 2010). This approach resembles the principle behind historical dictionaries, such as The Oxford English Dictionary, where the

meaning assigned to a specific word or notion is the conglomerate of its use through time, from the earliest recordings to the present (Simpson, 2013; cf. section 1.2).

In the analyses of information, misinformation, and disinformation from the perspective of 'meaning' the distinction between semantic theories and foundational theories of meaning is essential. The different kinds of theories of meaning answer different questions, wherefore, the adherence to a specific theory or notion of meaning has certain implications. For example, as Grice's theory of meaning is a foundational and mentalist theory, it entails questions and answers as to how groups or persons ascribe meaning to symbols and not what specific words or sentences mean. Thus, the function of the theory is to explain how meaning comes about and not why *S* means *p*. In other words Grice's theory is a pragmatic theory of meaning concerned with the use of language in communication. This does not mean that a given theory which employs Grice's notions of meaning (e.g. Dretske's theory of information) has to directly address questions of how meaning is ascribed, rather it is this 'level' of meaning which is present in the theory. It means that Grice's theory cannot be used to ask questions about what specific words and sentences mean. If that is the concern or interest, then another theory of meaning must be chosen.

3.2 Objective information and the rise of meaning

"In the beginning there was information. The word came later." (Dretske, 1981, p. vii). Thus begins *Knowledge and the Flow of Information* (Dretske, 1981). The quote is the briefest illustration of Dretske's conception of information – a commodity which pre-exists in the world independent of human agents, meaning, and language and which, by its mere existence, cannot be anything but true.

"It is common to think of information as a much later arrival on the evolutionary scene, as something that depends on the interpretive efforts – and, hence, prior existence – of intelligent life. (...) This is one way of thinking about information. It rests on a confusion, the confusion of *information* with *meaning*. Once this distinction is clearly understood, one is free to think about information (though not meaning) as an objective commodity, something whose generation, transmission, and reception do not require or in any way presuppose interpretive processes. (...) Meaning, and the constellation of mental attitudes that exhibit it, are manufactured products. The raw material is information." (Dretske, 1981, p. vii).

Information is thus a signal, which, qua its origin from natural sources – i.e. objects, events, states of affairs – is true par definition. In this sense truth is prior to meaning, however, as Dretske notes "Meaning is fine. You can't have truth without it." (Dretske, 2008, p. 29). This sounds like a paradox, when information defined by truth is prior to meaning, but it just

means that agents cannot assess whether or not something is true (e.g. talk about truth) without an understanding of the meaning (i.e. know the content) of that ‘something’. However, as information is an objective commodity independent of agents it does not pose a problem to the definition that agents cannot assess the information before meaning has entered the picture. How meaning enters the picture and enables information to ground knowledge is part of Dretske’s project in *Knowledge and the Flow of Information* (1981) and *Explaining Behavior* (1988).

According to Fred Adams (2015, personal communication) Dretske is concerned with how minds can rise out of purely physical events (i.e. information). The properties which minds rely on – which purely physical events do not have – are meaning⁶⁴, belief, and knowledge. In other words, Dretske is interested in how meaning and thereby mental concepts can rise out of a purely physical space – the nature. To answer this question Dretske introduces three orders of intentionality which a signal (structure, state, or event) can have.

In the 1st order of intentionality there is no lawlike relation between *F* and *G*:

“First Order of Intentionality

- (a) All *F*s are *G*
- (b) *S* has the content that *t* is *F*
- (c) *S* does not have the content that *t* is *G*

When this triad of statements is consistent, I shall say that *S* (some signal, event, or state) has a content exhibiting the first order of intentionality. So, for example, even though all Elmer’s children have the measles, and *S* carries the information (has the content) that *t* is one of Elmer’s children, *S* may not carry the information (have the content) that *t* has the measles.” (Dretske, 1981, pp. 172-173).

According to Dretske it is the 1st order of intentionality which information-processing systems exhibit. In the 2nd order of intentionality there is a lawlike or nomical relation between *F* and *G*:

“Second Order of Intentionality

- (a) It is a natural law that *F*s are *G*
- (b) *S* has the content that *t* is *F*
- (c) *S* does *not* have the content that *t* is *G*

⁶⁴ In Grice’s nonnatural sense.

When this triad is consistent, *S*'s content exhibits the second order of intentionality. For instance, one can believe (know) that the water is freezing without believing (knowing) that the water is expanding even though (let us say) it is nomically impossible for water to freeze without expanding – even though there is a natural law that tells us that water expands when it freezes.” (Dretske, 1981, p. 173).

In the 3rd order of intentionality there is an analytical or logical nestedness between *F* and *G*:

“Third Order of Intentionality

- (a) It is analytically necessary that *F*s be *G*
- (b) *S* has the content that *t* is *F*
- (c) *S* does *not* have the content that *t* is *G*

When this triad is consistent, *S*'s content exhibits the third order of intentionality. One might know (believe) that the solution to an equation is 23 without knowing (believing) that the solution is the cube root of 12,167. The fact that it is mathematically (analytically) impossible for 23 to be anything other than the cube root of 12,167 does not make it impossible for one to know (hence to be in a cognitive state having the content) that *t* = 23 without knowing (being in a cognitive state with the content) that *t* = the cube root of 12,167.” (Dretske, 1981, p. 173).

Cognitive structures such as meaning, belief, and knowledge exhibit the 3rd or at least the 2nd order of intentionality. The three orders of intentionality are one way to describe the idea of opacity – that is, the idea that beliefs and other cognitive states can represent one piece of information without representing the piece of information which is nested in the first piece of information. It is this idea of opacity, which Frege (1892) demonstrated with his ‘evening star’/‘morning star’ example. Here an agent can believe or know that ‘the evening star’ is a sense for the planet Venus without knowing or believing that ‘the morning star’ is another sense for the planet Venus. Therefore he can believe that the first identity statement (i.e. ‘evening star’ = Venus) is true and at the same time believe that the second identity statement (i.e. ‘morning star’ = Venus) is false although their truth conditions are the same as they share their reference (i.e. Venus) (cf. section 2.1.1).

The idea of nestedness (3rd order intentionality) as well as a cognitive system's ability to discriminate between the different pieces of information carried by a signal is very important for the possibility of the occurrence of meaning. Meaning arises when a cognitive structure (e.g. a particular belief) has some specific pieces of information as its propositional content without having other pieces of information as its propositional content – that is, when a structure's content exhibits the 3rd order of intentionality: “What we believe, and hence the beliefs themselves, must be distinguished even when their contents are interdependent. I will

call any propositional content exhibiting the third order of intentionality a *semantic content*.” (Dretske, 1981, p. 173).⁶⁵ The idea is that if G is logically nested in F then it follows that, if t is F , then t is also G . An example is water and H_2O . There is a nomical relation between water and H_2O , as water’s chemical formula is H_2O . This means that if a substance (t) is water (F), then the substance is also H_2O (G) – i.e. the information that t is F (that the substance is water) is also information about t being G (about the substance being H_2O). Thus, if a signal (or a structure) carries the information that t is F it also carries information about t being G , although it might not carry the information that t is G . “If a structure S carries the information that t is F , it does not necessarily carry the information that t is G even though nothing is F that is also not G . The information embodied in a structure defines a propositional content with intentional characteristics.” (Dretske, 1981, p. 172). For example, ‘ t is F ’ (the substance is water) and ‘ t is G ’ (the substance is H_2O) do not have the same meaning (sense) although they have the same reference. If an agent does not know about the relation between water and H_2O then it is possible for the agent to know or believe that a specific substance is water without knowing or believing that the substance is H_2O even though the signal carries information about the substance being both water and H_2O (Adams, 2015, personal communication; and Dretske, 1981).

Cognitive systems, for example human agents, are discrete when it comes to nomical, analytical, or logical nestedness which means that, as opposed to other kinds of systems, they can discriminate between the pieces of information nested in one another: “To occupy a belief state a system must somehow discriminate among the various pieces of information embodied in a physical structure and select *one* of these pieces for special treatment – as *the content* of that higher-order intentional state that is to be identified as the belief.” (Dretske, 1981, p. 174). This discrimination with respect to 3rd order intentionality gives rise to the different meanings (or different senses) between ‘the substance is water’ and ‘the substance is H_2O ’ and enables concept formation. In connection to a discussion about the concept of truth (cf. Ch. 4) Künne (2003) explains the idea of opacity (i.e. agents’ abilities to discriminate between different pieces of information) from the point of view, where a concept is already acquired by an agent:

“Having a concept is having a cognitive capacity: you have the concept of being thus-and-so, I take it, if and only if you are able to think of something *as* thus-and-so (or as

⁶⁵ Mind the difference between informational content and propositional or semantic content. The informational content of a structure always includes all the pieces of information nested in one another. A cognitive structure, i.e. a structure who exhibits the 3rd order of intentionality selects a specific piece of information as its only propositional or semantic content – that is, its meaning.

not thus-and-so). The concept of being F [e.g. salt] differs from the concept of being G [e.g. sodium chloride] if it is possible that somebody thinks of something as F without thinking of it as G . This may be the case even if (the property of) being F is (the property of) being G ." (Künne, 2003, p. 2).

Thus, on Dretske's account of information, meaning is the link between information as purely physical events in the world and mental capacities such as concept formation and cognition. This explains why Dretske holds that information and meaning must not be confused – information is that which enables meaning to enter the evolutionary scene. The discrimination between the different senses or pieces of information carried by a signal or a structure (i.e. its content) is dependent on the human agent's background knowledge. What information a signal or a cognitive structure carries depend on what the human agent already knows. If an agent knows the lawlike connection between F and G or knows that G is nested in F then the structure which carries the information that t is F also carries the information that t is G . This feature of background knowledge as partly determinant for what information a structure has as its content is what is encompassed by the k in Dretske's definition of informational content (cf. section 1.5.1):

INF_{DR} "*Informational content*: A signal r carries the information that s is F = The conditional probability of s 's being F , given r (and k), is 1 (but, given k alone, less than 1)" (Dretske, 1981, p. 65).

As mentioned, Dretske's ideas of opacity and 3rd order intentionality are very similar to Frege's idea of the difference between sense and reference. In *On Sense and Reference* (1892) Frege asks the question of what the meaning of a name, subsequently a sentence, is, whereas Dretske in *Knowledge and the flow of Information* (1981) asks the question of what the meaning of any symbol is (Adams, 2015, personal communication). However, Dretske (1981, 1983, 2008) does not refer to Frege (1892) for a notion of meaning. Instead Dretske refers to Grice's (1957) theory of meaning and his distinction between natural and nonnatural meaning (cf. section 2.2 and 3.2.1). This seems puzzling at first but the distinction between semantic theories and foundational theories of meaning might offer an explanation. The foundational theories try to explain how natural language acquires semantic properties, whereas the semantic theories explain these semantic properties. In the same way Dretske tries to explain how meaning and minds can rise out of physical space – that is how signals acquire their meaning. Thus, Dretske's scope is equivalent to the scope of the foundational theories of meaning – the difference is that Dretske wishes to explain meaning acquisition for all kinds of signals, not just linguistic ones. In this connection, it is important to note that semantic theories and foundational theories of meaning are not mutually exclusive. A given semantic

theory might put constraints on a given foundational theory, and vice versa, but both kinds of theories are needed as they answer different questions (Speaks, 2010).

3.2.1 Natural and nonnatural meaning

The reference to Grice and his distinction between natural and nonnatural meaning appears in *Knowledge and the Flow of Information* (1981), where Dretske, in a footnote, writes that his conception of information resembles or is equivalent to the notion of natural meaning in Grice's sense:

"There is, of course, a use of the term "meaning" that is close (if not equivalent) to the ordinary sense of "information". When we say, for example, that a blown fuse *means* the circuit has been overloaded, a sputtering engine *means* we are out of gas, and George's fingerprints on the glass *mean* that he was on the scene, we are using "meaning" in what Paul Grice calls its *natural* sense. When I say that information must be distinguished from meaning, I have in mind Grice's *nonnatural meaning*, the sense of meaning that is relevant to language and semantic studies." (Dretske, 1981, p. 242, footnote 10).

In this quote Dretske draws a distinction between Grice's two notions of meaning and take natural meaning to concern states of affairs in the actual physical world, and takes nonnatural meaning to concern language.⁶⁶ In Grice's theory meaning is reduced to the communicative intentions of agents (cf. section 2.2), however, in connection to the three orders of intentionality, Dretske describes meaning in terms of propositional content – semantic content – without reference to agents and their intentions. Moreover, in the description of the difference between information and meaning (in the nonnatural sense) only word- and sentence-meaning – i.e. semantic meaning – is present:

"Margaret's utterance, "I am alone" *means* that she is alone (that is what the words she uttered mean) even though her utterance may fail to carry this piece of information (e.g., when she is not alone). And when her utterance *does* carry the information that she is alone, it must also carry the information that she is *not* sipping martinis with Henry Kissinger. Yet, though her utterance necessarily carries the second piece of information whenever it carries the first, her utterance means that

⁶⁶ In *Meaning* (1957) both natural meaning and nonnatural meaning concern language. Grice even includes agents in certain forms of natural meaning: "I propose, for convenience, also to include under the head of natural senses of "mean" such senses of "mean" as may be exemplified in sentences of the pattern "A means (meant) *to do so-and-so* (by x)," where A is a human agent." (Grice, 1957, pp. 214-215).

she is alone without meaning that she is not sipping martinis with Henry Kissinger.” (Dretske, 1981, pp. 174-175).⁶⁷

Dretske’s definition of information as natural meaning in Grice’s sense, is further established in *Misrepresentation* (1986) and *Explaining Behavior* (1988), where Dretske prefers the notions ‘indication’ (i.e. has the function of indicating) and ‘natural meaning’ (in Grice’s sense) instead of information. “I have elsewhere (1981, 1983⁶⁸), under the rubric *information*, tried to say something more systematic about the idea of an objective, mind-independent, indicator relation. (...) Talking about information is yet a third way of talking about the fundamentally important relation of indication or natural meaning.” (Dretske, 1988, pp. 58-59). Besides the brief mention above, where Dretske establishes that information, indication, and natural meaning is essentially the same, he abandons the notion of information and sticks to indication and natural meaning.

3.2.2 Misrepresentation

In *Epistemology and Information* (2008) Dretske returns to information as the key term. He describes information as natural meaning in Grice’s sense and stresses the equivalence between natural meaning, indication, and information: “If *e* (some event) means, in the natural sense, that *s* is *F*, however, then *s* has to be *F*. Natural meaning is what indicators indicate. It is what natural signs are signs of. Natural meaning is information. It has to be true.” (Dretske, 2008, p. 31). According to Fred Adams (2015, personal communication) information, on Dretske’s account, is essentially relations between properties and representations of properties. These properties and the relations between the properties and their representations exist in the world independent of agents – independent of whether or not any agents get the information and understand it. Meaning (in the nonnatural sense) is ascribed to information in the process of belief- and concept formation. In other words, agents ascribe meaning to the information in order to understand and use it. However, the information itself is not dependent upon its meaning – i.e. its nonnatural meaning. The formation processes take place within the individual agents but the meaning ascription can be based on testimony by other agents, for example via books, pictures, oral testimonies, etc.

⁶⁷ In connection to how beliefs are structured Dretske does mention communicative intentions of speakers in connection to nonnatural meaning: “the fact that certain utterances, gestures, and signs have what Grice calls a nonnatural meaning (something with *semantic* content) derives from the semantic structure of those internal states (in particular, the intentions and beliefs) that give rise to that behavior. (...) our verbal behavior *means* something (in Grice’s nonnatural sense of “meaning”) because it is an established way of satisfying certain communicative intentions by agents with certain beliefs.” (Dretske, 1981, p. 204).

⁶⁸ *Knowledge and the Flow of Information* (1981) and *Précis of Knowledge and the Flow of Information* (1983).

When meaning has arisen and beliefs are being formed misrepresentation can occur. That is the reason why Dretske can hold that misinformation, which must be the result of misrepresentation, is not a kind of information. Information (true par definition) is prior to misinformation which is false. Misrepresentation, which is the source of misinformation (and disinformation⁶⁹), occurs when the informational link between an information source and a type (e.g. a semantic structure) is not preserved in the token (e.g. a particular belief) of the type:

“We need information to manufacture meaning (the concept) because information is required to crystallize a type of structure with the appropriate semantic content. (...) But once we have meaning, once the subject has articulated a structure that is selectively sensitive to the information about the *F*-ness of things, instances of this structure, tokens of this type, can be triggered by signals that *lack* the appropriate piece of information. When this occurs, the subject *believes* that *s* is *F* but, because this token of the structure type was not produced by the information that *s* is *F*, the subject does not *know* that *s* is *F*. And if, in fact, *s* is not *F*, the subject falsely believes that *s* is *F*. We have a case of misrepresentation – a token of a structure with a false content. We have, in a word, meaning without truth.” (Dretske, 1981, pp. 194-195).

When this ‘meaning without truth’ (i.e. nonnatural meaning without truth) is disseminated, received, or perceived by an agent it is misinformation. Thus, Dretske’s description of misrepresentation allows the following definition of misinformation:

MIS_{DR} Misinformation is semantic content that is false.⁷⁰

For example, if red light on a white object makes an agent perceive the object as red it is a case of misrepresentation. In this case there is no information that the object is red, the object is misrepresented as red and the agent is misinformed that it is red. The agent’s belief that the object is red is false as the object is actually white (Dretske, 1981).

⁶⁹ Disinformation is added to the decoy duck-example in Dretske (1983): “False information, misinformation, and (grimace!) disinformation are not varieties of information – any more than a decoy duck is a kind of duck. A glance at the dictionary reveals that information is related to intelligence, news, instruction, and knowledge – things that have an important connection to *truth*.” (Dretske, 1983, p. 57). Dretske does not define disinformation, but the addition of ‘grimace!’ suggests an understanding of disinformation in accordance to the definitions from the dictionaries (cf. section 1.2). Furthermore, Dretske does not address what the distinction between false information and misinformation is.

⁷⁰ Dretske (1981) mentions misinformation as that which is false. The definition of misinformation is derived from this statement and the description of misrepresentation as the process where falsity enters the picture.

3.3 Meaning in the Infosphere

Dretske's views on information as an objective semantic notion which is truthful and prior to meaning and which exists in the world independently of interpretive processes and intelligent life are greatly influential in the philosophy of information. Thus, the idea which Dretske expresses with the sentences "In the beginning there was information. The word came later." (Dretske, 1981, p. vii; cf. section 3.2) is adopted by Floridi and the Π Research Network (2013) and developed into the notion of the Infosphere in the maximalist interpretation of philosophy of information (in Floridi's interpretation of the field; cf. section 1.3.1). In the maximalist interpretation the universe is the infosphere, that is, the universe and everything in it is made of information in a 'facts-about-events' interpretation of information (as it is based on Dretske's 'environmental information' as Floridi (2005b) calls it). Thus, it follows that, in Floridi's interpretation, philosophy of information is "the philosophy of everything." (The Π Research Network, 2013, p. 33). The infosphere as a semantic environment (Floridi, 2002b) is part of what could be called the semanticization of the universe. It concerns how information as that which constructs the universe has acquired its meaning – i.e. the semanticization of information. As Floridi defines semantic information as

INF_{FL} Well-formed, meaningful, and truthful data

the semanticization process (and the problem) is moved from information to data: "How data can come to have an assigned meaning and function in a semiotic system in the first place is one of the hardest problems in semantics." (Floridi, 2005a, p. 358). The Π Research network describes the process through the reading of an altimeter: "When the altimeter responds to external changes in its environment it is not actually generating any information *per se*. What is generated might be called data. We need a syntactical structure – what philosophy of information calls a Level of Abstraction (...) – to give that data meaning and determine its informational content." (The Π Research Network, 2013, pp. 84-85). Thus, the data acquires meaning by inference and interpretation through a specific syntax, i.e. level of abstraction. The Π research Network (2013) develops this explanation on the basis of Floridi's (2005a) description of SDI⁷¹, however, Floridi (2005a) does not offer an explanation of this kind. Instead, Floridi (2011b) develops a new approach to the problem with semanticization of data. "It is based on a new theory of meaning – which I shall call *Action-based Semantics* (AbS)" (Floridi, 2011b, p. 162) – where meaning is the performed action prior to any

⁷¹ SDI: The standard Definition of Information. The Π Research Network (2013) refers to Floridi (2005a) where he uses the abbreviation SDI, however, the network use the abbreviation GDI for General definition of Information.

“semantic interaction with other agents” (Floridi, 2011b, p. 165). The technical details of AbS will not be entertained here as these are not necessary for the scope of exploring the connections between the notions of information, misinformation, and disinformation. The importance of the theory lies in the fact that it is a strictly semantic theory, wherefore, the requirement for ‘meaningfulness’ in the definition of information is a strictly semantic requirement. Therefore, semantic information can be a strictly semantic notion independent of informees. This view is different from the view held by Dretske, where semantic information is independent of any agents (Floridi, 2005a). As misinformation on Floridi’s (2011b) account is defined as

MIS_{FL(2011b)} “Well-formed and meaningful data (i.e. semantic content) that is false” (Floridi, 2011b, p. 260)

this is a strictly semantic notion as well, wherefore, it must also be independent of ‘informees’ (or perhaps misinformees). Misinformation is the direct opposite of information. Both information and misinformation are semantic content (i.e. well-formed, meaningful data) and they are divided by truth (information) and falsity (misinformation). Thus, on Floridi’s (2005a, 2011b) account, the connection or interrelation between information and misinformation differs from the connection or interrelation between the two notions on Dretske’s (1981) account, where information is prior to misinformation instead of its opposite. The connections between information and misinformation on the two different accounts are determined by ‘meaning’. On Floridi’s account semantic meaning in the form ‘meaningfulness’ is part of the definitions of information and misinformation, thus, they can be opposites in terms of truth-values. In contrast, meaning is not part of the definition of information on Dretske’s account. Information is prior to meaning (in the nonnatural, yet semantic sense, which resembles Floridi’s notion of meaningfulness) and misrepresentation is defined as ‘meaning without truth’, hence information and misinformation are divided both in terms of truth-values and in terms of whether or not meaning is part of their definitions.

As Floridi (2011b) defines disinformation in terms of misinformation (i.e. as a subset of misinformation) disinformation seems to be a strictly semantic notion independent of ‘informees’ (or disinformees) as well. However, when Floridi (2011b) defines disinformation he explicitly addresses the existence of an ‘informee’ (disinformee) in terms of a receiver:

DIS_{FL(2011b)} “‘Disinformation’ is simply misinformation purposefully conveyed to mislead the receiver into believing that it is information.” (Floridi, 2011b, p. 260).

Furthermore, an active agent is introduced in terms of the ‘purposeful conveying in order to mislead’.

3.4 Information and misinformation as communication?

How meaning and information are connected is at the heart of many theories of information and nearly as many different proposals have been offered (e.g. Budd, 2011; Dretske, 1981; Floridi, 2004b, 2005a, 2007, 2011b; Fox, 1983; Furner, 2004; Mai, 2013; and Scarantino & Piccinini, 2010). For instance, Budd (2011), Mai (2013), and Scarantino and Piccinini (2010) define information in terms of meaning. Scarantino and Piccinini develop two alethic neutral notions of information based on Grice's distinction between natural and nonnatural meaning (cf. section 4.2.3), whereas Budd places meaning⁷² as a constitutional part of information on the same level as truth. The truth condition is derived from Dretske (1981) to whom Budd refers when he argues that lies and deception are not information. This reference seems arbitrary as Budd argues that "*Information* is a meaningful communicative action that aims at truth claims and conditions." (Budd, 2011, p. 70). It is an agent-dependent definition of information as it presupposes agents to perform the communicative actions intentionally (cf. Ch. 5). Budd further argues that "[t]hose who would study information must begin with the understanding that information is a particular product of human invention." (Budd, 2011, p. 71) thus cementing the agent-dependency of information.

The communicative aspect of information is also the frame of Fox's (1983) theory of information and misinformation. Fox's theory is cast in terms of language – sentences and propositions. According to Fox "[i]nformation is not meaning, but meaning is evidently closely connected with information." (Fox, 1983, p. 95). 'Meaning' does not have the features which Fox ascribe to information – 'Meaning' cannot be asserted, it cannot be the bearer of truth-values, and it cannot be the object of propositional attitudes. Propositions have all these features and they "are atemporal and non-spatial; they are abstract objects, like sets, numbers, or sentence types." (Fox, 1983, p. 78). Thus, Fox develops a propositional analysis of information, where information is reduced to propositions and defined in term of the information carried by sentences:

INF_{FO} "The *information* carried by a set \$ of sentences is the proposition p such that

1. p is the conglomerate proposition expressed by a set \$' of sentences appropriately associated with \$,
2. there exists an individual X, the originator of \$, who is in a position to know that P." (Fox, 1983, p. 200).

⁷² Budd considers meaning as a semantic concept although he recognizes that indexicality and thereby context and historicity are important features (Budd, 2011).

The formulation of ‘set of sentences’ and ‘conglomerate proposition’ accounts for information as a mass noun which has collectiveness and divisiveness – that is, many pieces of information can be collected to one new piece of information, and one piece of information can be divided into many pieces of information. Furthermore, information, on this account, need not be true and it need not be believed by anyone, neither informer nor informee (Fox, 1983). Fox develops a similar propositional analysis of misinformation, where the only difference is, that misinformation is false:

MIS_{FO} “The *misinformation* carried by a set \$ of sentences is the proposition p such that

1. p is the conglomerate proposition expressed by a set \$’ of sentences appropriately associated with \$,
2. there exists an individual X, the originator of \$, who is in a position to know that P,
3. p is false.” (Fox, 1983, p. 201).

On these accounts of information and misinformation ‘meaning’ enters the picture

“as the crucial mediating factor between sentences and propositions by virtue of which the former are able to express the latter. A sentence is a string of symbols meeting certain conditions of grammaticality, and by itself it does not express any claim about the way the world is. A proposition, on the other hand, is a language independent representation of the way the world is, with no essential connection to sentences whatever. Meaning is the bridge that connects sentences with propositions, enabling the former to express the latter.” (Fox, 1983, p. 96).

Exactly how this ‘bridging’ works is left unsolved but it is essentially the same problem as the problem of semanticization which Dretske (1981) and Floridi (2005a, 2011b) face. Although propositions in themselves are language independent the definitions of information and misinformation are cast in terms of propositions expressed by sentences, wherefore these two notions are dependent upon or bound by language. Information and misinformation are what is communicated – “Given the propositional analysis of information, the consequence of all this is that the information carried by a sentence, or a set of sentences, is *determined* by meaning, and hence *relative* to meaning.” (Fox, 1983, p. 97). The proposition which constitutes the information cannot be expressed without the use of meaning and Fox introduces three different notions of meaning – a semantic notion and two pragmatic notions. Thus, what information a sentence carries is determined by what that sentence literally means, what it means on some specific occasion, or what some agent means by that sentence:

“Typically, what is understood to be the *meaning* of a sentence (i.e., the meaning₀) is very different from what a *person* means in using the sentence (the meaning₁) and from what the *sentence* means (the meaning₂), on some occasion. Hence in discussing “the meaning of a sentence” it is necessary that we clearly understand whether it is the “basic” or “non-contextual” notion of meaning (meaning₀) or one of the “extended” or “contextual” notions of meaning (meaning₁ or meaning₂), which is under discussion.” (Fox, 1983, p. 83).

Information is for Fox (1983) bound by language that is further established by his analyses of the acts of informing and misinforming, which are conducted in order to analyze the notions of information and misinformation. According to Fox (1983) informing (and misinforming) is a kind of telling⁷³ and telling requires hearing and understanding on the part of the one who is told something. Thus, informing requires hearing and understanding on the part of the informee as well. In order to define the relation between informing and information Fox develops a Linking Principle (LP):

LP “X informs Y that P if X passes the information that P to Y.” (Fox, 1983, p. 190),

which is a 3-place relation between X, P, and Y. Fox also develops a Complementary Linking Principle (CLP) for misinforming and misinformation:

CLP “X misinforms Y that P if X passes the misinformation that P to Y.” (Fox, 1983, p. 191).

The Linking Principle is based on what Fox calls a truism and is supposed to be a precise specification of the truism:

Truism “Information is that which is conveyed when X informs Y that P.” (Fox, 1983, p. 190).

However, in the “truism” the relation between informing and information seems to be the other way around from the Linking Principle: in LP informing takes place when information is passed on, in the truism information is conveyed when informing takes place. Taken together, the “truism” and LP says that informing takes place when information is passed on (LP) and when informing takes place information is conveyed (“truism”). On Fox’s account both

⁷³ Fox defines informing as: “X *informs* Y that P only if X tells Y that P.” (Fox, 1983, p. 116) and he defines telling as: “X *tells* Y that P if X says to Y that P and Y hears and understands what X says, namely, that P.” (Fox, 1983, p. 114).

informing and information are bound by language and communication. Therefore, it does not make a difference which way the link between informing and information is stated. Through the notion of telling, hearing and understanding on the part of the informee are built into the act of informing. Therefore, when information is that which is conveyed in the act of informing (the “truism”) it follows that in order for information to be conveyed it requires that the informee hears and understands it. The same follows from LP when the mere providing of information establishes the act of informing which requires hearing and understanding. That ‘information conveyed’ requires hearing and understanding is fairly uncontroversial as it follows from Fox’s contain/convey distinction. ‘Convey’ is used to denote that which someone actually learns from some sentences, utterances, or communication. If a hearer does not understand what is being said, or does not come to learn anything, then the utterances do not convey any information. However, the utterances might contain a lot of information which is not conveyed to the hearer because he does not learn anything from the utterances (Fox, 1983)⁷⁴.

In contrast, on Floridi’s (2005a, 2005b) account of information, which is not bound by language and communication, the two statements (the “truism” and LP) are not two instances of the same principle. Floridi would agree that the act of informing requires the provision of information in some way or other. However, it follows from his view – the view that information can be present even though we cannot access it – that information cannot solely be that which is conveyed through the act of informing. Thus, Floridi would accept LP but he would not accept the “truism” as a definition of information. On Floridi’s account information is independent of informees – i.e. information is objective and independent of whether or not any informee hears or understands it. A sentence written in Spanish still carries information even though the reader does not understand Spanish. Dretske’s (1981) definition of information as an objective commodity does not allow for the “truism” either. Thus, the framework within which a theory of information, misinformation, and disinformation is cast determines which notions can be developed. A theory cast within the framework of communication (Fox) allows for different conclusions than a theory developed to explain cognition, concept formation, and learning (Dretske) and different conclusions than a theory developed as part of an ontological program cast within an engineering framework (Floridi; cf. sections 4.2.2 and 4.2.5). This is why Carnap (1950) specifies that the fruitfulness of an explicatum must be evaluated within or according to a specified context – i.e. within or according to the context or framework within which it is developed. However, the shared

⁷⁴ How this fits with the “truism” which seems to imply that information is only ‘information conveyed’, is not specified.

context of automatic detection enables some sort of comparison or evaluation of the theories against one another.

3.5 Disinformation

Meaning determines what is told, what someone is informed of, and which information is passed on and conveyed on Fox's account. What is told – i.e. what is heard and understood – is determined by occasion specific sentence-meaning(₂), speaker-meaning(₁), and the invariant sort of meaning(₀) – i.e. by the semantics and the pragmatics of the utterance with a distinction between literal meaning and speaker-/occasion meaning. Misinformation, which is a kind of information (i.e. the subset of information that is known to be false) is determined by meaning in the same way. Thus, information and misinformation are essentially the same – propositions expressed through sentences determined by meaning – the only difference being the specified truth-value (false) for misinformation. Fox's notions of information and misinformation fit, in terms of similarity, as explicata of the explicanda from Vocabulary.com, where information can be a message received and understood and misinformation is incorrect information (cf. section 1.2). The account is fairly simple and exact due to the formal formulation of information and misinformation (i.e. INF_{FO} and MIS_{FO}) within the context of communication. Furthermore, within a context of communication an account where information and misinformation is that which is communicated is fruitful and captures the intuitions of everyday use of the notions. It also seems to be a fruitful account for automatic detection as it is already cast in terms of language and take both semantic and pragmatic features of 'meaning' into account. However, when the notion of disinformation enters the picture it gets more complicated. Fox (1983) has not developed an account of disinformation, wherefore such an account must be found elsewhere. The most extensive conceptual analysis and well-argued views on disinformation is Fallis' (2009, 2011, 2014, 2015) accounts (cf. section 1.5).

Fallis' (2014) account of disinformation and, indirectly, misinformation and information, is somewhat broader than Fox's accounts of information and misinformation. On Fallis' account disinformation, misinformation, and information are not limited to being kinds of telling. Based on the distinctions between lying, misleading, and deceiving Fallis argues that "unlike lies, disinformation does not have to be a statement. Fetzer's [2004b] analysis incorrectly rules out what we might call *visual disinformation*." (Fallis, 2014, p. 138). Visual disinformation can be forged maps, doctored photographs, and the like. Instead Fallis (2014) defines disinformation as

$DIS_{FA(2014)}$

disinformation is information that is intentionally misleading (Fallis, 2014).

Fallis defines information as

$INF_{FA(2014)}$ information is all that with representational content (Fallis, 2014)

wherefore it follows that disinformation is all that with representational content which is intentionally misleading. Fallis' definition of information and his definition of misinformation as

$MIS_{FA(2014)}$ misinformation is inaccurate information (Fallis, 2014)

further establish that telling is neither sufficient nor necessary for informing, misinforming, and disinforming to take place. However, it is possible to disinform by telling, but the act of disinforming is broadened to capture showing as a means to disinforming as well (Fallis, 2014). Showing and telling as means to disinforming are based on Grice's notion of nonnatural meaning. In *Meaning* (1957) Grice uses examples of showing and telling in order to establish and explain the three intentions which must be present in communication with nonnatural meaning. In order for nonnatural meaning to be present the mere showing or telling is not enough. The intentions behind the showing or telling must be to convince someone (an audience) about what is shown or told. Moreover, it must be the showing or telling itself along with the recognition (on the part of the audience) of the intention which convince the audience about what is shown or told – i.e. to disinform them on Fallis' account (cf. section 2.2). Fallis (2014) uses Grice's examples of showing and telling in order to derive a classification of different varieties of disinformation.

3.5.1 Content and meaning

In his various analyses of disinformation Fallis does not explicitly address which notion of meaning is presupposed by 'representational content'. The notion of representational content is equivalent to Floridi's notion of semantic content as well-formed, meaningful data, although representational content is broader as it captures pictures, gestures, and the like (Fallis, 2015, personal communication). Thus, at least a semantic notion of meaning is presupposed by Fallis' notion of information as representational content. Moreover, as it is the content which has the function to mislead or must be intentionally misleading in order for disinformation to occur (and must be inaccurate and unintended misleading in order for misinformation to occur) the need for 'meaning' is further established:

"What if a piece of information is misleading (that is, it is likely to create false beliefs), but is not misleading about the accuracy of its content? For instance, suppose that you create a map of South America so that it looks like it was drawn by Europeans in the seventeenth century (for example, by its being on old parchment, with ornate lettering) to try to misle[a]d people into thinking that Machu Picchu was discovered

in the seventeenth century. This is certainly information that has the function of misleading. Thus your fake map counts as disinformation in my new analysis. However, the *content* of this map is completely accurate; it is not going to mislead anyone about the location of Machu Picchu. With this sort of case in mind, we might want to amend my new analysis and reserve the term *disinformation* for information whose *representational content* has the function of misleading.” (Fallis, 2015, pp. 418-419).⁷⁵

When it is the representational content which has the function of misleading it is at least the literal meaning of the content which causes the misleading for example if the content is false. However, when the definitions of misleading and deceiving are recalled a pragmatic notion of meaning enters the picture as well. Misleading and deceiving are defined in terms of Gricean implicatures (false implicatures) where it is the implicatum which is misleading or deceiving (cf. section 1.1). The implicatures are generated because the speaker exploits the maxims of the Cooperative Principle whereby the literal meaning of what is said is extended by the speaker-meaning (i.e. pragmatic ‘meaning’). Thus the distinction between word- and sentence-meaning and utterer’s meaning is enforced as instances of nonnatural meaning. The use of false implicatures enables Fallis (2014) to derive a notion of ‘true disinformation’. The notion of ‘true disinformation’ might seem contradictory especially when the explicanda⁷⁶ of disinformation are considered (cf. section 1.2) and when it is the representational content which has the function to mislead. However, the implicature arise due to a violation of the supermaxim of quantity⁷⁷ as well as the supermaxim of manner⁷⁸ and the false implicatum counts as part of the content. Thus, with true disinformation it is still the content which is misleading. The same holds for misinformation, in Fallis’ terms, as it is the inaccuracy of the content which renders it misleading (Fallis, 2009, 2011, 2014, 2015). Thus, due to Fallis’ adherence to Grice’s notion of implicature as well as showing and telling as instances of nonnatural meaning – used as a supplement to the literal content of pictures and utterances – both semantic and pragmatic notions of meaning are at play.

The entrance of true disinformation renders the connections between the notions of information, misinformation, and disinformation more complex. On Dretske’s and Floridi’s accounts true disinformation is impossible. On Fox’s account true disinformation also seems

⁷⁵ Even though the map might not count as disinformation due to the accuracy of the content it is still misleading and therefore still counts as deception (cf. section 1.1).

⁷⁶ Disinformation is often defined in terms of falsity and at least inaccuracy.

⁷⁷ “Make your contribution as informative as required.” (Grice, 1989, p. 26).

⁷⁸ “Be perspicuous.” (Grice, 1989, p. 27).

to be impossible as he defines misinformation in terms of falsity and disinformation often is defined as a subset of misinformation. Therefore, in order to further investigate and analyze the connections between information, misinformation, and disinformation the notion of truth as the second asymmetry must be taken into account.

3.6 Meaning and truth

The various accounts of information, misinformation, and disinformation examined employs different notions of meaning. Some accounts refer to a strictly semantic notion of meaning, whereas others leave room for pragmatic meaning as well. These different notions of meaning are linked to the notions of information, misinformation, and disinformation in different ways. Thus, on Dretske's (1981) account of information defined as natural meaning in Grice's sense nonnatural meaning, in a purely semantic sense, is ascribed to information by agents in order to make sense of the informational content of signals. 'Meaning' is determinant for misinformation as the result of misrepresentation as misinformation is 'meaning without truth'. Therefore, information is prior to misinformation, which is not a direct opposite to information. On Floridi's (2005a, 2005b, 2011b) account information and misinformation are direct opposites. They are both determined by semantic meaning and truth-value. Together they comprise the set of semantic content, i.e. well-formed, meaningful data – information is the inherently truthful subset, whereas misinformation is the false subset. Disinformation is the subset of misinformation which is deliberately false.

The notions of information and misinformation are cast in terms of communication on Fox's (1983) account. 'Meaning' is not an explicit part of the definitions of information and misinformation, however, 'meaning' is the link between propositions and sentences and information and misinformation are defined as propositions expressed through sentences. Therefore, meaning is determinant for information on Fox's account. As the notions are cast in terms of communication both semantic meaning in terms of sentence-meaning and pragmatic meaning in terms of speaker-meaning and occasion-meaning are present. For Fox information and misinformation are not opposites. Misinformation is a kind of information, that is, the 'known-to-be-false' kind, where information itself is alethically neutral. The notion of meaning is not explicated on Fallis' (2009, 2011, 2014, 2015) account of disinformation either. However, as disinformation can be based on Gricean implicatures and is construed intentionally by disinformers pragmatic meaning in terms of speaker-meaning must be present along with semantic meaning. Gricean implicatures are generated due to pragmatic features of meaning where the speaker-meaning goes beyond word- and sentence-meaning. On Fallis' account disinformation and misinformation are not opposite to information. Information is defined as alethically neutral, misinformation is defined as inaccurate and

misleading information, and disinformation as intentionally misleading information which can be both true and false.

In short, the accounts of information where truth is a requirement only deals with semantic meaning as part of information, misinformation, and disinformation (Floridi, 2005a, 2005b, 2011b) or in connection to information and as a part of misinformation (Dretske, 1981, 1983, 2008). The accounts where information is alethically neutral and defective information is the outset (e.g. disinformation; Fallis, 2009, 2011, 2014, 2015), or at least equal to information in the analyses (e.g. misinformation; Fox, 1983), pragmatic meaning is present as well.

As mentioned, due to the disagreement about the truth requirement for information and Fallis' introduction of true disinformation as an actual possibility, the further investigation of the interconnections of information, misinformation, and disinformation has to be conducted through the aspect of 'truth' as the second asymmetry.

Ch. 4 Truth

4 Truth

Questions about truth – such as what truth is, how to define it, its ontological status, its relation to knowledge, etc. – have been present in philosophy ever since the ancient Greeks. Truth as a condition for knowledge put forth by Plato in *Theaetetos* in, what is known as, the standard definition of knowledge – Justified True Belief (JTB) – naturally raises the questions of what truth is and how to secure it. These questions seek (and possibly find) their answers within different branches of philosophy:

“It is important to be clear about what is meant by a theory of truth. A theory of truth is a theory of what truth is or, alternatively, of what something’s being true consists in. It is not a theory of how truth is arrived at or of what the criterion for success is on that score; such questions belong to epistemology, whereas questions about the nature of truth are part of metaphysics and the philosophy of language.” (Collin & Guldmann, 2005, p. 89).

The debates about the different theories of truth, as elucidated by Künné (2003), first concern whether truth is a property and, if that is the case, which kind of property and second what the fundamental truth-bearers are. For instance, if truth is a relational property, is it then a relation towards other truth-value bearers (Coherence theory), a relation towards objects (Object-Correspondence), a relation towards events (Event-Correspondence), or is truth identical with facts (Identity Theory)? These questions remain neutral as to what the fundamental truth-bearers are. Two options present themselves as truth-bearers (or as that which truth is a ‘property of’ in Künné’s terminology): sentences and propositions. If truth is a property of sentences, is it then of type-sentences or of token-sentences, and can it be explained and finitely stated? If truth is a property of propositions, then the questions arise

whether the property is stable, whether it can be explained, and whether it can be finitely stated? (Künne, 2003). Those who accept propositions as truth-bearers usually also accept sentences as derivative truth-bearers in the sense that a sentence is true because the proposition it expresses is true. In contrast, those who accept sentences as the fundamental truth-bearers, do so because they deny the existence of propositions (Smid, 2015, personal communication). Thus, it is possible to hold that sentences and propositions are both truth-bearers iff (if and only if) propositions are the fundamental truth-bearers.

The various debates will not be entertained in detail here as the scope is neither to settle on one theory of truth as ‘the theory of truth’, nor to discuss the concept of truth in itself. Rather, the scope is twofold: First, the scope is to single out which notions of truth might be at play in those definitions of information where truth is a requirement; As the requirement for truth is not uniform across all definitions of information – that is, some definitions of information do not have truth as a requirement – the second part of the scope is to discuss the differences between those definitions of information that do require truth and those definitions that do not require truth. These discussions of information and truth encompass discussions of the differences between those definitions of misinformation and disinformation, respectively, where falsity is a requirement, and those definitions where falsity is not a requirement. In other words, is information necessarily true or is it alethically neutral and are misinformation and disinformation, respectively, necessarily false or are they alethically neutral? However, some of the main theories of truth will be briefly outlined in order to be able to satisfy the first part of the scope.

4.1 Theories of truth

Within philosophy of information the notions of truth and falsity are applied in connection to discussions, theories, and analyses of information, misinformation, and disinformation. However, the notions of truth and falsity themselves are hardly (if ever) discussed and it is rarely specified which theory of truth is presupposed when truth is a requirement for information. Some suggestions have been made as to which theories of truth might be adhered to within philosophy of information. The suggestions regard some of the most common theories of truth and enclose the deflationary theory, correspondence theory, and the semantic conception of truth, which, therefore, will be presented.

4.1.1 Correspondence

The correspondence theory of truth – or rather, the correspondence theories of truth, as there have been formulated many varieties – is the most common theory of truth. In general correspondence theory is the view that truth is a relational property between a truth-bearer and a truth-maker. The variations, differences, and disputes between the various conceptions

of correspondence theories concern what counts as a truth-bearer and especially what counts as a truth-maker – i.e. ‘that something’, which the truth-bearer corresponds to (David, 2002; Künne, 2003).

A truth-bearer is that entity, which has, or can be ascribed, a truth-value (either true or false) within a given theory of truth. Although it says ‘truth’ in truth-bearer and truth-value it refers to both truth and falsity. Thus ‘truth-bearer’ is an entity capable of holding truth or lacking truth (i.e. be false), that is, it has a ‘truth-value’, which can be positive (true) or negative (false). When something (e.g. a truth-bearer) has ‘a truth-value’ it just means that it is either true or false. A truth-bearer can only have one specific truth-value (i.e. true or false), however, the truth-value needs not be specified; it can be un-known. Therefore, it is perfectly legitimate to say that a sentence, for instance, has a truth-value even though it is un-known whether the sentence is true or false⁷⁹. A truth-bearer can also be that “of which it makes sense to ask whether they are true or false, thus allowing for the possibility that some of them might be neither.” (David, 2002, section 2.1). This means that a truth-bearer need not have a truth-value as, for example, in the case of sentences stated in imperative, i.e. ‘Close the door’ or in interrogative, i.e. ‘Is the door closed?’. As examples of different candidates for truth-bearers David (2002) mentions “beliefs, thoughts, ideas, judgments, statements, assertions, utterances, sentences, and propositions.” (section 2.1). David acknowledges, in line with Künne (2003), that sentences and propositions are the most common candidates as truth-bearers.

Truth-bearers and truth-values do not do the work alone. There must be something in the world which the truth-bearers correspond to (or fail to correspond to) in order to hold a truth-value. The correspondence is to a truth-maker. “A truthmaker is anything that makes some truthbearer true. Different versions of the correspondence theory will have different, and often competing, views about what sort of items true truthbearers correspond to (facts, states of affairs, events, things, tropes, etc.).” (David, 2002, section 2.2). Thus, it is the correspondence between some specified truth-bearer (e.g. proposition) and a specified truth-maker (e.g. state of affairs) which determines whether the truth-bearer is true or false (i.e.

⁷⁹ This is the case when the truth conditions can be specified but cannot be satisfied. That is, it is possible to specify under which circumstances the sentence would be true (i.e. which conditions must be met in order for the truth-bearer to be true), but either these circumstances does not obtain or it is un-known whether they will obtain (or have obtained). This is the reason why Frege (1892) holds that sentences which include fictitious names do not have references (truth-values) – the truth conditions can be specified but they cannot be satisfied. For example, the truth conditions for the sentence ‘Zeus is a Greek god’ are that ‘Zeus’ refers to an actual object which has the property of being a Greek god. However, as ‘Zeus’ is a fictitious name these truth conditions cannot be met wherefore the sentence does not have a truth-value (cf. section 2.1.1).

determines the truth-value). The most common candidates as truth-makers are facts (David, 2002; Künné, 2003; Collin & Guldman, 2005) or alternatively states of affairs (David, 2002). This means that in most versions of correspondence theory it is either facts or states of affairs that truth-bearers have to correspond to.

For a correspondence theorist the challenge lies in the description of what it means for some entity to correspond to something in the world, as well as the description of what it is, which has to correspond (Collin & Guldman, 2005) – that is, of all the different possibilities, what is the fundamental truth-bearer, what is the truth-maker, and how should the relation between the two be specified?

4.1.2 Semantic conception of truth

The semantic conception of truth is developed by Tarski and proposes an extensional definition of truth. The extensionality consists in the idea that all the things which fall under a given concept (e.g. truth) can be listed as part of that concept which then can be finitely stated. As sentences are the essential truth-bearers (i.e. truth is a property of sentences) in Tarski's definition, such a list would consist in all true sentences (Collin & Guldman, 2005). However, the truth-values are unknown for many sentences regarding the past, the future, and much in between. Tarski proposed that sentential truth can be explained and that the explanation consists in the description of the truth-conditions for any actual or potential sentence within a specified language (Collin & Guldman, 2005; Künné, 2003). Thus, the definition of truth contain a sentence and the condition under which that sentence is true, for instance,

“Tarski was a logician’ is true if and only if Tarski was a logician.’

Tarski's theory of truth encounters problems when applied to natural languages where sentences can be self-referential and can ascribe truth and falsity to themselves. In order for the semantic theory to work “the truth definition for a given language must always be formulated in another language. (...) it follows from this, according to Tarski, that no general definition of truth can be given. Truth can only be defined for a concretely specified, precisely regimented language.” (Collin & Guldman, 2005, p. 109). Thus, truth can only be defined for formal languages (e.g. predicate logic and other logics) and mathematics. As a response to this constraint disquotational or redundancy theories of truth have been offered (Collin & Guldman, 2005; Künné, 2003).

4.1.3 Deflationism

The redundancy theory of truth is also known as the deflationary theory of truth:

“The deflationary theory has gone by many different names, including at least the following: the redundancy theory, the disappearance theory, the no-truth theory, the disquotational theory, and the minimalist theory. There is no terminological consensus about how to use these labels: sometimes they are used interchangeably; sometimes they are used to mark distinctions between different versions of the same general view.” (Stoljar & Damnjanovic, 1997, introduction).

Hence, in Collin and Guldman (2005) the view is presented as redundancy theory with special emphasis on “a “disquotational” form of the theory” (p. 115). Künne (2003), on the other hand, simply avoids the term ‘deflationism’ with the argument that it has been applied to such various fields as nihilism, disquotationalism, and minimalism. In Künne’s descriptions of ‘truth’ and theories of truth, nihilism is what one gets if truth is not a property, disquotationalism is what one gets if truth is a property of sentences which can be explained but cannot be finitely stated, whereas minimalism is what one gets if truth is a property of propositions which can be explained but cannot be finitely stated.

The general view, which according to Stoljar and Damnjanovic (1997) (as well as Collin and Guldman, 2005), is captured under the heading deflationism etc. is that

“to assert that a statement is true is just to assert the statement itself. For example, to say that ‘snow is white’ is true, or that it is true that snow is white, is equivalent to saying simply that snow is white, and this, according to the deflationary theory, is all that can be said significantly about the truth of ‘snow is white’.” (Stoljar & Damnjanovic, 1997, introduction).⁸⁰

The deflationary theory of truth with all its different varieties – or more precisely the various different theories comprised under a single heading – “has been one of the most popular approaches to truth in the twentieth century” (Stoljar & Damnjanovic, 1997, section 1) and it is suggested that “the view originates with Frege.” (Stoljar & Damnjanovic, 1997, section 1).

⁸⁰ In the literature on ‘assertion’ three different accounts of assertion are debated – the knowledge-account, the belief-account, and the truth-account of assertion. On the knowledge-account, to assert p entails that one knows p; on the belief-account, to assert p entails that one believes p; and on the truth-account, to assert p entails that p is true (Brown & Cappelen, 2011). On the knowledge-account and the truth-account of assertion the assertion of p is in itself a statement regarding the truth of p (as truth is constitutive of knowledge). On the belief-account of assertion the truth of p is not entailed by the mere assertion of p.

However, the connection between deflationism and Frege is debated. Given the connection between Frege's theory of sense and reference and Dretske's theory of information some additional views on Frege's conception of truth will be taken into account.

4.1.4 Propositional primitivism and identity theory

In contrast to Stoljar and Damnjanovic (1997), Künne (2003) argues that "[p]ropositional primitivism is an important ingredient in Frege's reflections on truth" (Künne, 2003, p. 16) where propositional primitivism is the view that truth is a property of propositions and that propositional truth cannot be explained (Künne, 2003). Frege's notion of 'thought' is equivalent to 'proposition' i.e. the sense of a declarative sentence which has a truth-value as its reference (cf. section 2.1.1), wherefore it is propositions which are the fundamental truth-bearers on Frege's account. Moreover, Frege's conception of truth has also been identified with the identity theory of truth, where truth is a relational property identical with facts, as he has assented to the identity thesis which states that

“(*Idem*) For all x, x is a true proposition iff (if and only if) there is a fact with which x is identical.” (Künne, 2003, p. 6).

However, Künne (2003) argues that “by assenting to (*Idem*), Frege did not mean to explain the concept of truth. Rather, he took (*Idem*) to explain the notion of fact.” (Künne, 2003, p. 16). Derived from Frege's definitions of the thought (proposition) as the sense of a declarative sentence and of the relation between sense and reference, Künne offers an explanation for why facts and true propositions are not the same as it is otherwise claimed in the identity theory of truth:

“At any rate, the next observation should carry conviction: we individuate facts less finely than true propositions. The fact that you never met Cassius Clay is the same as the fact that you never met Muhammad Ali, the fact that I am German is identical with the fact that WK [Wolfgang Künne] is German, the fact that three-quarters of the electorate went to the polls is the same as the fact [that] 75 per cent of the electorate went to the polls, and the fact that for cooking we often need some common salt is identical with the fact that for cooking we often need sodium chloride. But (as Frege would be the first to insist) in each of these cases my utterances of the embedded sentences express different propositions.” (Künne, 2003, pp. 11-12).

These differences between facts and propositions, where 'fact' – as a state of affairs – in some sense is prior to 'proposition' – which has different meaning ascriptions – bear resemblance to Dretske's ideas of information as state of affairs, signals, and events, where meaning is ascribed subsequently (cf. section 3.2). These connections between Dretske, Frege, and propositional primitivism suggest that the latter – although it is not one of the common

theories of truth – have to be taken into account in the further discussions of information, misinformation, and disinformation – along with correspondence theory, the semantic conception of truth and deflationism.

4.2 Truth, falsity, and alethic neutrality

In order to explore the asymmetry between information, misinformation, and disinformation the discussions of ‘meaning’ (cf. Ch. 3) do not provide a complete picture. In addition, it is necessary to discuss how truth (and falsity) plays into the various definitions of information, misinformation, and disinformation. As mentioned, these discussions entertain a twofold scope: to identify which theory of truth is adhered to when truth is a requirement; and to discuss the requirement for truth/falsity vs. the requirement for alethic neutrality. As it is a twofold scope and not two separate scopes the various discussions will be intertwined. Thus, analyses of which theories of truth are present within specific definitions of information will be conducted as part of the discussions of the requirement for truth/falsity vs. alethic neutrality for information, misinformation, and disinformation, respectively.

4.2.1 False information and ‘false information’

As discussed in Ch. 3, there is no consensus as to how to define semantic information, neither in philosophy of information nor outside the field and especially not between the different fields which comprise philosophy of information⁸¹. The definition of information differs between the different fields in accordance to its use – i.e. the notion differs in accordance to what it is needed for. This mechanism of ‘definition according to need’ is especially clear when the explicanda of misinformation and disinformation⁸² are taken into account (cf. sections 1.2.1-1.2.2). In order to be able to define misinformation and disinformation as clear and simple as possible a notion of false information is needed, wherefore references to wrong or false information are given without any disclaimers that there is no such thing as false information. Given the goals of the detecting-projects (i.e. to algorithmically detect information, misinformation, and disinformation in online social network structures) and the possibility of discrepancies between what the different tools detect (due to differences in their definitions of information, misinformation, and disinformation; cf. Introduction) the notion of false information and whether or not such information is possible is crucial.

⁸¹ E.g. computer science, cognitive science, artificial intelligence, philosophy, information studies, etc.

⁸² That is, the definitions of misinformation and disinformation from the dictionaries.

In Floridi's conception of philosophy of information 'false information' is a contradiction in terms which has to be understood in a specific sense – that is, 'false' denotes that the content has nothing to do with the original notion – 'false information' is not a specific kind of information (Floridi, 2005a, 2007). Floridi (2011b) explicitly states this point when he argues that "[o]ne speaks of false information in the same way as one qualifies someone as a false friend, i.e. not a friend at all." (Floridi, 2011b, p. 183) which is opposed to, for example, 'false teeth'⁸³ which are a kind of teeth and 'false proposition'⁸⁴ which is still a proposition independent of its truth-value. Floridi derives these conclusions about the application of 'false' from analyses of *predicative* vs. *attributive* uses of adjectives:

"Take two adjectives like "male" and "good". A male constable is a person who is both male and employed as a policeman. A good constable, however, is not a good person who is also employed as a member of the police force, but rather a person who performs all the duties of a constable well." (Floridi, 2005a, p. 364).

Whether an adjective is used predicatively or attributively can be determined through a splitting test. If the adjective can be split from the compound of which it is a part without semantic loss, then the adjective is used predicatively and if the adjective cannot be split from the compound, then it is used attributively (Floridi, 2005a). Thus in the example of 'male constable' and 'good constable' 'male' is used predicatively because 'male' applies to the person in question independently of his being a constable, whereas 'good' is used attributively because it applies directly to his property of being a constable – the two cannot be split. Along the same lines 'false' can be used either predicatively or attributively. When negative adjectives (such as 'false') are used attributively "they negate one or more of the qualities necessary for x to be x" (Floridi, 2005a, p. 365) as in the example of the 'false friend' who is not a friend. According to Floridi (2005a) 'false' in 'false' proposition is used predicatively because it can be split into '*p* is a proposition' and '*p* is false' without semantic loss, whereas 'false' in 'false information' is used attributively because it cannot be split into '*p* is information' and '*p* is false': "it would be a mistake or an act of misinformation to assert, that *p* constitutes information about the number of natural satellites orbiting around the earth *and* is also a falsehood." (Floridi, 2005a, p. 365). However, this conclusion presupposes a notion of information as inherently truthful (as well as a notion of misinformation as necessarily false and different in kind from information) – i.e. it presupposes the exact notion of information, which Floridi argues in favor of. That is, a notion where information is constituted by truth

⁸³ Example from Floridi (2007).

⁸⁴ Example from Scarantino and Piccinini (2010), which is also used in Floridi (2011b).

and is not a truth-bearer as truth-bearers are always capable of holding either of the truth-values ‘true’ and ‘false’(cf. section 4.1.1)⁸⁵:

“According to [SI]⁸⁶, semantic information is, strictly speaking, inherently *truth-constituted* and not a contingent *truth-bearer*, exactly like knowledge but unlike propositions or beliefs, for example, which are what they are independently of their truth-values and then, because of their truth-aptness, may be further qualified alethically.” (Floridi, 2011b, pp. 183-184).

However, if it is not accepted that information requires truth in order to be information, then ‘false information’ does pass the splitting test (i.e. ‘false’ is used predicatively). For instance, on Fallis’ account of information as representational content ‘false information’ would be split into ‘*p* is representational content’ and ‘*p* is false’ which is roughly equivalent to the example with false propositions. On Fox’s (1983) account, where information is reduced to propositions, ‘false’ is also applied predicatively to information. Thus, Floridi’s (2005a) splitting test and the distinction between predicative and attributive applications of adjectives do not provide an argument as to why information is inherently truthful. The same line of reasoning is pursued by Scarantino and Piccinini (2010) when they discuss whether ‘false information’ is analogous to ‘false policeman’ or analogous to ‘false proposition’. They argue in favor of alethic neutrality for information and conclude that

“[w]hether false information passes the splitting test depends on whether we accept that a false *p* can constitute nonnatural information. We do! And so, we submit, do most other people, including most cognitive scientists and computer scientists. As far as we are concerned, *that the earth has two moons* is both nonnatural information and false” (Scarantino & Piccinini, 2010, p. 321).

Fox (1983) also argues in favor of information as alethically neutral, and thereby accepts false information as a kind of information. Through extensive analyses of various sentences that include the terms ‘inform’, ‘misinform’, ‘information’, and ‘misinformation’ Fox – who seems to be the sole researcher who has tried to develop a philosophical account of misinformation

⁸⁵ It follows from the description of a truth-bearer as that entity which is capable of holding truth-values that one cannot claim information to be a truth-bearer and at the same time claim that it can only hold one of the truth-values (i.e. truth).

⁸⁶ [SI] is in square-brackets in the original. The definition of [SI], Semantic Information, reads “[SI] *p* qualifies as semantic information if and only if *p* is (constituted by) *well-formed, meaningful, and truthful data*.” (Floridi, 2011b, p. 183). [SI] is a summary of GDI* which is Floridi’s (2011b) notation for the revised general definition of information. For an outline of Floridi’s abbreviations, cf. Ch. 1, note 30.

(and this in relation to an account of information) – concludes that information need not be true, whereas misinformation must be false:

“All the evidence is in favor of the thesis that misinformation must be false. Since information may be false, we see that misinformation is a species of information, just as misinforming is a species of informing. Thus a coherent picture of the relationships between informing, information, misinforming, misinformation, and truth has emerged: informing does not require truth, and information need not be true; but misinforming requires falsehood, and misinformation must be false.” (Fox, 1983, p. 193).⁸⁷

Fox’s conclusion is based on two questions of entailment:

- “1. Do sentences of the form ‘X informs Y that P’ entail P?
2. Do sentences of the form ‘X misinforms Y that P’ entail not-P?”

(Fox, 1983, p. 160).

Fox’s answers to 1. and 2., which lead to the conclusion, are that in relation to the first question “the entailment in question does *not* obtain; informing does not require truth. However, the answer to the second question is just the opposite: the entailment in question *does* obtain; misinforming does require falsity.” (Fox, 1983, p. 160). However, these questions of entailment only provide answers for the actions ‘informing’ and ‘misinforming’ in connection to truth. Fox’s conclusions about ‘information’ and ‘misinformation’ in connection to truth are derived from the Linking Principle⁸⁸ for ‘inform’ (i.e. the action) and ‘information’ (i.e. the entity) as well as the Complementary Linking Principle⁸⁹ for ‘misinform’ (i.e. the action) and ‘misinformation’ (i.e. the entity).⁹⁰ When the entailment in question 1. does not obtain it follows that the truth-value of information, which is passed on through the act of informing, is un-specified. Contrary, when the entailment in question 2. does obtain it follows that the truth-value of misinformation, which is passed on through the act of misinforming, is specified – the misinforming ‘that P’ entails not-P, i.e. the truth-value of the misinformation ‘that P’ is false as P is false.

⁸⁷ This ‘coherent picture’ resembles the connections of the explicanda from Vocabulary.com as well as the account put forth by Fallis (2015) – except that Fallis does not require falsity for misinformation (or for disinformation).

⁸⁸ The Linking Principle: “X informs Y that P if X passes the information that P to Y.” (Fox, 1983, p. 190).

⁸⁹ The Complementary Linking Principle: “X misinforms Y that P if X passes the misinformation that P to Y.” (Fox, 1983, p. 191).

⁹⁰ For a discussion of the Linking Principle see section 3.4.

Fallis (2009, 2011, 2014, 2015) agrees with Fox that ‘information’ does not require truth and he agrees that misinformation (and disinformation) is a kind of information. However, Fallis does not agree that misinformation requires falsity per se. Instead Fallis (2009, 2014, 2015) defines misinformation in terms of inaccuracy and un-intended misleadingness.

Disinformation, according to Fallis, does not require falsity either as the notion must be able to capture a true variety – i.e. true disinformation (Fallis, 2009, 2011, 2014, 2015). The true variety of disinformation is needed in order for the notion to be in line with the definitions of ‘deceiving’ and ‘misleading’ which are present within the definition of disinformation (cf. sections 1.1-1.2). As seen in Ch. 1 and 2 (sections 1.1 and 2.3.3) ‘misleading’ and ‘deceiving’ can be performed through false implicatures (i.e. implicatures where what is literally said is true but what is implicated is false) therefore, when disinformation is defined in terms of ‘misleading’ and ‘deceiving,’ this possibility must be present for disinformation as well.

4.2.2 Semantic information as inherently truthful

Although Floridi (2005b) adopts Fox’s (1983) definition of misinformation, he is not convinced by Fox’s ‘coherent picture’ (cf. section 4.2.1) where misinformation is a kind of information. Floridi endorses “the Dretske-Grice approach: meaningful and well-formed data constitute semantic information only if they also qualify as contingently truthful.” (Floridi, 2005a, p. 351) (cf. sections 1.5.1 and 3.3). Floridi’s definition of semantic information is generated as the result of an analysis of the Standard Definition of Information (SDI):

“SDI) σ is an instance of DOS information if and only if:

SDI.1) σ consists of *n data* (*d*), for $n \geq 1$;

SDI.2) the data are *well-formed* (*wfd*);

SDI.3) the *wfd* are *meaningful* ($\text{mwfd} = \sigma$).” (Floridi, 2005a, p. 353)⁹¹.

Condition (SDI.1) comes with two comments: first, SDI.1 entails that information cannot be dataless (i.e. SDI endorses typological neutrality); second, SDI.1 entails that information can consist of a single datum δ and SDI endorses both taxonomic neutrality (“a datum is a relational entity”, Floridi, 2005a, p. 357) and ontological neutrality (“no information without data representation”, Floridi, 2005a, p. 357). Condition (SDI.3) comes with one comment:

⁹¹ “the symbol σ and the term “infor” [are used] to refer to discrete items of information, irrespective of their semiotic code and physical implementation.” (Floridi, 2005a, p. 353).

“How data can come to have an assigned meaning and function in a semiotic system in the first place is one of the hardest problems in semantics. (...), the semanticisation of data need not detain us here because SDI.3 only requires the δ to be provided with a semantics already. The point is not *how* but *whether* data constituting semantic information can be correctly described as being meaningful *independently* of an informee. The *genetic neutrality* (GN) supported by SDI states that:

GN) δ can have semantics *independently* of any informee.” (Floridi, 2005a, p. 358).⁹²

According to Floridi SDI.1-SDI.3 and the endorsement of the four neutralities are necessary conditions for semantic information. However, when SDI.1-SDI.3 are taken to be jointly sufficient it follows that SDI also endorses alethic neutrality (AN) (cf. section 1.5.2), which entails that

FI) false information (including contradictions), i.e. misinformation, is a genuine type of DOS information, not pseudo-information;

TA) tautologies qualify as information; and

TI) “it is true that σ ” where σ is a variable that can be replaced by any instance of genuine DOS information, is not a redundant expression; for example, “it is true” in the conjunction “‘the earth is round’ qualifies as information and it is true” cannot be eliminated without semantic loss.”

None of these consequences seems ultimately defensible, and their rejection forces a revision of AN and hence of SDI.” (Floridi, 2005a, pp. 359-360).

Floridi then deals with false information through a consideration of, what he calls, “Nine bad reasons to think that false information is a type of semantic information” (Floridi, 2005a, p. 360) as well as two positive arguments (e.g. attributive vs. predicative uses of ‘false’, cf. section 4.2.1) in favor of the conception that ‘false information’ is ‘pseudo-information’ (i.e. not a type of information) – he leaves the two other entailments (TA and TI) for later treatment. Each ‘bad reason’ is met by an objection in favor of the view that semantic information requires truth. However, most of the objections fall short of the same critique as Floridi’s argument for the attributive use of ‘false’ in connection to information (cf. section

⁹² The solution for the problem with semanticization of data is offered in Floridi (2011b) – i.e. six years after the current proposal for a definition of semantic information. However, the proposed solution does not influence the argumentation for semantic information as well-formed, meaningful, and truthful data, which is the same in Floridi (2011b) as in Floridi (2005a).

4.2.1). For instance, Floridi's objection to reason no. 5 (FI.5) presupposes that 'genuine information' equals 'truthful information':

"FI.5) FI is meaningful and has the same logical structure as genuine information.

Objection: this is simply misleading. Consider the following FI: "One day we shall discover the biggest of all natural numbers". Being necessarily false, this can hardly qualify as genuine but false information. It can only provide some genuine, non-primary information μ , e.g. about the mathematical naivety of the source." (Floridi, 2005a, p. 362).

What count as genuine information, depends on the definition of information that is endorsed. If SDI is taken to be sufficient for semantic information, then 'genuine but false information' does not pose a problem. Even the reason itself, as stated by Floridi (i.e. FI.5), presupposes a difference between 'false information' and 'genuine information', i.e. that genuine information is truthful. The 'the same as' construction entails, in this case, that 'false information' and 'genuine information' are two separate entities, which are compared. Furthermore, it is questionable whether Floridi's objection is actually an objection to FI.5. In Floridi's framework 'false information' or misinformation (as these are equivalent to Floridi) is defined as well-formed and meaningful data that is false (Floridi, 2005a). Thus, the meaningfulness (in the semantic sense) of 'false information' is not questioned in the objection. However, the objection does not explain why 'false information' fails to have the same logical structure as genuine information – which must be the part of FI.5 that is objected to – the objection simply states that something false cannot be genuine information.

From the nine objections to 'false information' as a kind of information and the two positive arguments in favor of 'false information' as pseudo-information (1. The attributive vs. predicative uses of 'false', cf. section 4.2.1, and 2. That the inclusion of 'false information' as semantic information makes it impossible to account for semantic erosion or semantic loss⁹³) Floridi concludes that

⁹³ Scarantino and Piccinini (2010) provide an objection to Floridi's argument of the unaccountability of semantic loss when 'false information' is included in semantic information (or that the conversion of true semantic content into its negation does not result in semantic loss). According to Scarantino and Piccinini there are different kinds of semantic loss, for instance, quantitative and qualitative. They agree that on the quantitative level the negation of true statements in some domain does not provide any semantic loss. However, qualitatively there is a loss, an epistemic loss as falsehoods are less epistemically valuable than truths (Scarantino & Piccinini, 2010).

“Well-formed and meaningful data may be of poor quality. Data that are incorrect (...), imprecise (...) or inaccurate (...) are still data and they are often recoverable, but, if they are not truthful, they can only constitute misinformation (let me repeat: they can be informative, but only indirectly or derivatively, for example they can be informative about the unreliability of the source,...). We have seen that misinformation (false information) has turned out to be not a type of information but rather pseudo-information. This is the Dretske-Grice approach (...):

[...] *false* information and *mis*-information are not kinds of information – any more than decoy ducks and rubber ducks are kinds of ducks. Dretske (1981), 45.

False information is not an inferior kind of information; it just is not information. Grice (1989), 371.” (Floridi, 2005a, pp. 365-366).

Thus, Floridi revises SDI into RSDI, which includes a necessary condition for truth (in terms of truthfulness):

“RSDI) σ is an instance of DOS information if and only if:

1. σ consists of n data (d), for $n \geq 1$;
2. the data are *well-formed* (wfd);
3. the wfd are *meaningful* ($mwfd = \delta$);
4. the δ are *truthful*.”

(Floridi, 2005a, p. 366).

Besides the requirement for truth Floridi’s definition of semantic information does not have much in common with Dretske’s definition of semantic information (cf. section 3.2). For Dretske ‘truth’ is secured by Grice’s notion of natural meaning as equivalent to information (Dretske, 1981, 2008). Information is properties of actual states of affairs, events, and signals with occur in the physical world, wherefore it cannot be anything but true. Dretske’s requirement for truth is formulated relative to a probability distribution. For something (e.g. s is F) to be information s has to raise the probability of F to 1. Thus, Dretske’s theory of semantic information is a probabilistic theory with regard to informational content. However, as Dretske’s theory is based on Shannon’s mathematical theory of communication it is also probabilistic with regard to the amount of information carried by a signal. The amount of information is determined by probabilities in the way that the more likely some event is to occur at the source of the signal, the less information that signal carries about the source and vice versa. This level of probability in regard to amount of information enables Dretske to hold that tautologies and other necessary truths do not generate any information. Their likelihood of occurrence is 1, wherefore the amount of information carried by the signal is 0, despite their candidacy for being information in terms of a probability of 1 in terms of their

content (i.e. that s is F). However, “although there is no information associated with water’s being H_2O (or, indeed, any necessary truth), the belief that water is H_2O constitutes an assimilation of two *different* cognitive structures in terms of their *sameness* of content.” (Dretske, 1981, p. 218). Thus, within the cognitive system the belief that a specific substance is water is connected with the belief that the same substance is H_2O . The assimilation does not generate a new piece of information, however, it does generate a new insight for the cognitive system in question.

4.2.3 Grice in disguise – an objection to Dretske and Floridi

Scarantino and Piccinini (2010) offer an account of information as a two-fold notion based on Grice’s notions of natural and nonnatural meaning, thus arriving at natural and nonnatural information, where nonnatural information is roughly equivalent to representation⁹⁴. “(...) we explicitly distinguish between a notion of information that can help naturalize intentionality (natural information) and a notion of information that itself needs to be naturalized (nonnatural information).” (Scarantino & Piccinini, 2010, p. 315). Scarantino and Piccinini offer their account within philosophy of information from the intersection between cognitive science, computer science, and philosophy of language. The naturalization of intentionality is to explain how cognition arises and thus Scarantino and Piccinini (2010) have the same project as Dretske (1981). The notions of information follow Grice’s (1957) notions of meaning in the way that what is natural meaning on Grice’s account is natural information on Scarantino and Piccinini’s account and what is nonnatural meaning on Grice’s account is nonnatural information on Scarantino and Piccinini’s account. In this way specific spots are natural information about measles, whereas three rings on the bell in a bus are nonnatural information about the bus being full. The difference between Grice and Scarantino and Piccinini is that Scarantino and Piccinini cast their account in terms of probabilities which suggest a Dretskeian reading of Grice (in terms of probabilities about s being F and o being G , etc.). However, probabilistic accounts of natural and nonnatural information enable them to argue against the Veridicality Thesis for both natural and nonnatural information. That natural information does not require truth as part of its definition is controversial within

⁹⁴ “We agree that non-natural information is in need of naturalistic explication just as much as intentional content. “To carry non-natural information about” and “to represent” are (for present purposes) synonymous expressions, and representation is precisely what a naturalistic theory of intentionality aims to explicate.” (Scarantino & Piccinini, 2010, p. 315).

Floridi's conception of philosophy of information⁹⁵ – especially when the notion is based on natural meaning in the Gricean sense which entails truth – and it is in direct conflict with Dretske's (1981) account of semantic information⁹⁶, which Scarantino and Piccinini argue against. On Scarantino and Piccinini's view, natural information is about raising the probability of something being the case (the same holds for nonnatural information) even though this 'something' might not be the case. In this way the spots are natural information about measles because the occurrence of those exact spots on a person raises the probability that the person has measles (although it is not information *that* measles are present; cf. section 3.2). The probability is not necessarily raised to 1 as there might be cases where a person with those spots does not have the measles (Scarantino & Piccinini, 2010). The notion of natural information thereby allows for the possibility of error something which is not possible on Grice's account of natural meaning (Grice, 1957) and which is not possible on Dretske's account of semantic information as equivalent to natural meaning (Dretske, 1981, 1983, 2008). That the probability is not necessarily raised to 1 is the main difference between Scarantino and Piccinini's (2010) account of natural information and Dretske's (1981) account of semantic information. Dretske's account is also cast in terms of probabilities (as it is based on Shannon's mathematical theory of communication), however, Dretske requires that the probability is raised to 1 in order for a signal to carry information about its source.

Even though Scarantino and Piccinini's definition of natural information allows for error "[t]alk of false information applies primarily to (...) nonnatural information." (Scarantino & Piccinini, 2010, p. 319) as the mistakes which might be made on the basis of natural information are due to imperfect correlations between internal states and external stimuli. Thus, the natural information which is acted upon is present, however, the event which it is information about might fail to obtain (Scarantino & Piccinini, 2010).

Scarantino and Piccinini (2010) argue that if the Veridicality Thesis for nonnatural information (VT_{NN}) is adopted then "unbeknownst to most cognitive scientists, cognition also involves the processing of *misinformation*, where misinformation is understood as different in kind from information." (Scarantino & Piccinini, 2010, p. 324). Thus they reject Dretske's and Floridi's interpretation of misinformation as something opposite to information, and refrain from the use of 'misinformation' as a term in general. Instead Scarantino and Piccinini speak

⁹⁵ It is Floridi's (2005a, 2005b, 2007) account of semantic information as inherently truthful – and thereby his conception of philosophy of information – which Scarantino and Piccinini (2010) argue against when they reject the Veridicality Thesis for nonnatural information.

⁹⁶ It is Dretske's (1981) account of semantic information which Scarantino and Piccinini (2010) explicitly argue against when they reject the Veridicality Thesis for natural information.

of “false (nonnatural) information” (Scarantino & Piccinini, 2010, p. 319) and specify that “[t]he notion of nonnatural information used in cognitive science is best interpreted as the notion of *representation*, where a representation is by definition something that can get things wrong.” (Scarantino & Piccinini, 2010, p. 324). Information as representation which can be either true or false is equivalent to Fallis’ (2009, 2014, 2015) notion of information as representational content as well as Floridi’s (2003, 2004b, 2005a, 2005b, 2007, 2011b) notion of semantic content in general – i.e. the standard/general definition of information, which Floridi argues is not a sufficient definition of semantic information.

Scarantino and Piccinini’s notion of nonnatural information is an agent-dependent or agent-relative notion as it must presuppose agents with the three intentions that constitute nonnatural meaning. In line with Grice’s theory of meaning (cf. section 2.2) it must also be possible to speak of nonnatural information in convention based instances of nonnatural meaning such as traffic lights. Thus, the meaning_{NN} of the red light, that the cars have to stop, is nonnatural information that the cars have to stop. In this way, agents and their actions and intentions enter Scarantino and Piccinini’s account of nonnatural information.

4.2.4 Ordinary language use

The conception of agents and their actions and intentions are one of the differences between Fox’s (1983) and Floridi’s (2005a, 2007) accounts of information. Fox’s and Floridi’s analyses of, and argumentations for, whether or not information entails truth (cf. sections 4.2.1 – 4.2.2) are dependent on how they conceive of language – especially ordinary language use, i.e. what people actually say. Fox (1983) uses the fact that people quite unproblematic say ‘false information’ as argumentation for the claim that information does not entail truth. Floridi (2005a), on the other hand, argues that ‘false’ in ‘false information’ operates the same way as in ‘false policeman’ – i.e. to denote that the policeman (the information) is not a policeman (information) at all (cf. section 4.2.1).

Fetzer (2004a) argues along the same lines as Fox and states that “common uses of “misinformation” implies that *false information* is not an anomalous conception, which by itself raises doubts about any theory that information cannot be false.” (Fetzer, 2004a, p. 228). Misinformation, in Fetzer’s terms is false or misleading information. Fetzer’s argument about misinformation and false information is part of the conclusion of his inquiry which sets out to “demonstrate that an account of his [Floridi’s] kind is too narrow to be true and that its adoption would hopelessly obscure crucial differences between information, misinformation, and disinformation.” (Fetzer, 2004a, p. 223). On Fetzer’s view

“Floridi must shoulder a substantial burden to overcome the apparent benefits of the standard conception [of information, i.e. well-formed, meaningful data], however, not

least at all because there are three basic kinds of sentences – *exclamatory*, *imperative*, and *interrogatory* – that are typically considered to be “meaningful” but are not ordinarily considered to be true or false. It should therefore be borne in mind that his attention is focused upon *declarative sentences*, synthetic ones specifically, rather than upon sentences generally.” (Fetzer, 2004a, pp. 223-224).

Scarantino and Piccinini (2010) capture these kinds of sentences as their notion of nonnatural information need not be truth-evaluative. Furthermore, this notion of nonnatural information also captures the examples, that Fetzer (2004a) provides, of information as blood spots, maps, and the Rosetta Stone – i.e. information which is not expressed by sentences at all. The examples also make evidence for Fetzer’s claim that information is context-dependent: “The [blood] spot, the map [of the British Museum], and the [Rosetta] stone implicitly reflect the context-dependence of information, in the sense that how much information can be extracted from the data tends to depend upon the background, knowledge, and abilities of those who interact with it.” (Fetzer, 2004a, p. 227). Floridi (2005a and 2005b) pick up the example of the Rosetta Stone and uses it to argue in favor of his own account of information as an informee-independent account (cf. sections 3.3 and 4.2.2).

Fetzer (2004a) argues against Floridi’s (2004b) account of strongly semantic information in favor of a semiotic and pragmatic account of information where information can be false:

“The *semiotic conception*, according to which something qualifies as information when it stands for something for someone, makes information relative to someone for whom that something so stands. (...) And it coheres with a *pragmatic conception* of truth for which something is true when it provides guidance appropriate for the pursuit of our goals.” (Fetzer, 2004a, pp. 226-227).

Fetzer (2004a) argues that false information and tautologies are also information, two claims which Floridi (2004b, 2005a, 2007) rejects. Furthermore, Fetzer (2004a) argues that “it is true that...” is non-redundant, something which Floridi (2004b, 2005a, 2007) also rejects. Floridi (2005a, 2005b) argues that when information is true par definition, the mere assertion of a declarative sentence (which qualifies as information because it is true) in itself states its truth, whereby the addition “it is true that...” is redundant. This conception at first resembles the redundancy theory of truth (i.e./or the deflationary or disquotational theory, depending on one’s preferences and conception of truth) but is actually used by Floridi to argue against the deflationary theories of truth:

““It is true that” in “it is true that *p*” could be redundant in view of the fact that there cannot be factual information that is not true, but DTT [deflationary theories of truth] could mistake this linguistic or conceptual redundancy for unqualified dispensability.

“It is true that” could be redundant because, strictly speaking, information is not a truth-bearer but already encapsulates truth as truthfulness. Thus, DTT may be satisfactory as theories of truth-ascriptions while being inadequate as theories of truthfulness.” (Floridi, 2005b, section 3.2.3).

Thus, on Floridi’s account of information there is no truth-bearer neither sentences nor propositions. Truth is inherent in information in another way (cf. section 4.2.5).

In contrast, Fetzer’s (2004a) statement about names, definite descriptions, predicates, and sentential functions that “they obviously lack the sentential or propositional structure to be true or false” (Fetzer, 2004a, p. 229) suggests that he sees both propositions and sentences as truth-bearers – i.e. propositions as fundamental truth-bearers and sentences as derivative truth-bearers. Thus, Fetzer’s and Floridi’s very conceptions of truth and truth-ascription conflict in the same way as their definitions of information. And it is because of their conflicting conceptions of truth that they disagree about the definition of information. For Fetzer (2004a) truth supervenes on information whereas truth for Floridi (2004b, 2005, 2007) is encapsulated by information.

4.2.5 Truth and the Veridicality Thesis

The Veridicality Thesis is introduced in Floridi (2007) as a label for the conception that semantic information requires truth. Thus, if a philosopher or researcher endorses the veridicality thesis it means that this philosopher or researcher requires semantic information to be truthful par definition. The veridicality thesis does neither specify a specific theory of truth nor say anything about the nature of truth (i.e. what truth amount to) or how it is acquired.

In the article *In Defence of the Veridical Nature of Semantic Information* (2007) Floridi offers a logical treatment of the Bar-Hillel-Carnap Paradox. According to Floridi (2004b, 2007) the Bar-Hillel-Carnap Paradox is one of the main challenges to quantitative theories of semantic information and therefore one of the main reasons why the standard, or general, definition of semantic information (i.e. well-formed, meaningful data) needs revision. In Floridi’s (2004b) article the Bar-Hillel-Carnap Paradox serves as the outset for the description of – as well as the arguments for – semantic information as inherently truthful. The paradox is information-theoretic in nature and based on the way bits of information are calculated. Shannon’s Mathematical Theory of Communication (information-theory) deals exclusively with the ‘amount of information’ carried in a signal and not with the ‘content of information’ within a message – i.e. “the semantic aspects of communication” (Carnap & Bar-Hillel, 1952, p. 1). Therefore, when semantics are taken into account (i.e. information content) the paradox arises in connection to the amount of information carried in self-contradictory sentences. In

the Mathematical Theory of Communication the amount of information carried by some occurrence (e.g. 'heads' by the toss of a coin) is determined by the probability of that occurrence (i.e. the probability of 'heads' – 0.5). The more likely an occurrence is the less information is carried in that occurrence and vice versa. In other words, high probability (i.e. > 0.5) means a low amount of information, whereas low probability (< 0.5) means a high amount of information (Carnap & Bar-Hillel, 1952; Dretske, 1981). This means that contradictions (i.e. 'p and $\neg p$ '), which have zero probability, carry the greatest amount of information possible (Floridi, 2005b, 2007). The paradox is exactly this: that contradictions, which par definition are false (as their occurrence is impossible), carry the most information (Floridi, 2004b, 2005b, 2007). In a footnote Scarantino and Piccinini (2010) describe the paradox in terms of exclusion of possible worlds: "The paradox emerges if we follow Carnap and Bar-Hillel (1952) in holding that the information content of a sentence corresponds to the set of possible worlds excluded by its truth. On this assumption, contradictions have maximal information content, because they exclude all possible worlds." (Scarantino & Piccinini, 2010, footnote 18, p. 328). It is important to note that Bar-Hillel and Carnap did not themselves consider the paradox a problem per se⁹⁷

"It might perhaps, at first, seem strange that a self-contradictory sentence, hence one which no ideal receiver would accept, is regarded as carrying with it the most inclusive information. It should, however, be emphasized that semantic information is here not meant as implying truth. A false sentence which happens to say much is thereby highly informative in our sense. Whether the information it carries is true or false, scientifically valuable or not, and so forth, does not concern us. A self-contradictory sentence asserts too much; it is too informative to be true." (Carnap & Bar-Hillel, 1952, pp. 7-8).

Scarantino and Piccinini (2010) acknowledge that Floridi's (2004b, 2005a, 2005b, 2007) definition of semantic information solves the paradox. The solution is the requirement for truth in the definition of semantic information (i.e. an emphasis on alethic values rather than probability). A contradiction, which by definition states 'p and $\neg p$ ', cannot be truthful as a whole. When it is required that information is truthful the contradiction does not count as information, which means that it cannot provide the maximum amount of information. Thus, the problem is solved – contradictions do not carry the maximum amount of information as they are not truthful and therefore fail to be information in the first place (Floridi, 2005b, 2007).

⁹⁷ Thanks to Emmanuel Genot for emphasizing this point.

Floridi's (2004b, 2005a, 2005b, 2007, 2011b) explicit defense of the veridicality thesis as well as Scarantino and Piccinini's (2010) explicit rejection of it do not provide any insights as to which notion of truth is inherent in the theories that endorse the thesis. As mentioned above, the veridicality thesis does not come with a specified or fixed theory of truth. Such a theory is to be specified within each theory of information that endorses the veridicality thesis. If the label 'The Veridicality Thesis' is fully adopted it follows that Dretske (1981, 1983, 1988, 2008) and Grice (1989)⁹⁸ endorse it, whereas Fallis (2009, 2011, 2014, 2015), Fetzer (2004a), and Fox (1983) reject it.

According to Fred Adams (2015, personal communication) it is most often the disappearance theory of truth (cf. section 4.1.3 on deflationism) or the correspondence theory, which is presupposed in philosophy of information. This view is partly shared by Floridi who states that "in the literature on semantic information there appears to be at least an implicit predilection for some version of a Tarskian and/or correspondentist approach." (Floridi, 2011b, p. 184)⁹⁹. In a footnote to this quote Floridi refers to Dretske (1981) and Fox (1983) among others. However, according to Adams (2015, personal communication) Dretske (1981) employs the disappearance theory of truth, where 'truth' just means facts. This inconsistency in the application of different theories of truth to one and the same theory of information – in fact one and the same book – might stem from the disputes of how to label the different theories of truth as well as how to use the different labels (cf. section 4.1).

Floridi (2011b) is well aware that a theory of truth is needed in order to answer the question "What does it mean for semantic information to be truthful?" (Floridi, 2011b, p. 184) and he states that he will answer the request for a theory of truth. As mentioned (in section 1.3.1) Floridi's conception of truth is closely connected to his constructionist view of knowledge as well as the method of levels of abstraction (LoA). The constructionism guides his quest for a theory of truth as he sets out to 'reverse engineer' semantic information itself instead of applying a pre-given theory of truth to semantic information. The 'reverse engineering' approach "consists in assuming the artefact itself as given – that is, in assuming that we do have in our hands a piece of (truthful) semantic information – and then trying to discover the principles governing its properties and workings by analyzing its structure, function and operations." (Floridi, 2011b, p. 185). The 'reverse engineering' starts with a translation of

⁹⁸ Grice (1989) does not provide a theory of information. However, in his *Retrospective Epilogue* under Strand Six on the Conversational Maxims and the Cooperative Principle he states that "False information is not an inferior kind of information; it just is not information." (p. 371).

⁹⁹ Note that 'philosophy of information' and 'the literature on semantic information' are not necessarily completely overlapping.

non-propositional semantic information into propositional semantic information i – e.g. ‘The beer is in the fridge’¹⁰⁰. i can then be polarized into a query (Q) and a result (R) – e.g. ‘Where is the beer?’ + ‘in the fridge’. The query must be made within a context (C), at a specific LoA (L), and with a specific purpose (P) – the CLP parameters – and the result must be given within the same CLP parameters (Floridi, 2011b). After the polarization all the semantic content is shifted to the side of the query by normalizing it into a yes/no question ($Q_{0/1}^{CLP} + A_{0/1}$) – e.g. ‘Is the beer in the fridge?’ + ‘Yes/no’.

“[I]t is now easy to see that it is really Q and not A that sets the scope of the CLP parameters. A Boolean answer [0/1 – i.e. no/yes] can only endorse the *context* in which, the *level of abstraction* at which, and the *purpose* for which the Boolean question is formulated; it can neither change nor challenge them.” (Floridi, 2011b, p. 191).

Through polarization and normalization the semantic information i has been disassembled into two components:

$$i^{CLP} = Q_{0/1}^{CLP} + A_{0/1}$$

“On the one hand, we have seen that $Q_{0/1}$ sets the CLP parameters. Since it provides all the content in i , $Q_{0/1}$ also identifies its referent, that is, what i is about. We can express all this more precisely by saying that $Q_{0/1}^{CLP}$ identifies a system s (the referent of i) and provides all the *semantic content* (the content in i) for a *model* of s (namely $Q_{0/1}^{CLP} + A_{0/1}$) within a given context, at a particular LoA and for a purpose. (Floridi, 2011b, p. 193).

The next step is that the answer $A_{0/1}$ must saturate the question $Q_{0/1}^{CLP}$. Saturation consists in verification of the CLP parameters and validation of the system:

“ $A_{0/1}$ saturates $Q_{0/1}^{CLP}$ by implicitly *verifying* its CLP parameters (roughly: both ‘yes’ and ‘no’ implicitly signal that the question is being asked in the right context, at the right LoA and for the right purpose) and explicitly *validating* its content, as a model of the system (roughly: ‘yes’ and ‘no’ provide a green or a red light for the question respectively.” (Floridi, 2011b, p. 194).¹⁰¹

¹⁰⁰ The example is from Floridi (2011b).

¹⁰¹ Recall that the correctness theory of truth is explained as a case of reverse engineering, wherefore, the terminology is borrowed from software engineering (Floridi, 2011b) – hence, the references to saturation, verification, validation, systems, models, and red and green lights.

As Floridi's theory of truth is a correctness theory the next step is to define correctness. Correct validation consists in the generation of a model m of the system s , where the model m is a proxy such that if an agent a holds $Q_{0/1}^{CLP} + A_{0/1}$, then a 's proximal access to the model m commutes with a 's distal access to the system s . " a 's *proximal access* to m commutes with a 's *distal access* to s if and only if a can read/write s by reading/writing m ." (Floridi, 2011b, p. 197).¹⁰² Thus, the analysis of correctness enters "into the realm of pragmatics, that is, into the realm of actual and hopefully successful *interactions*" (Floridi, 2011b, p. 196) between the agent a , the model m , and the system s by "an explicit reference to an informee a [a holds $Q_{0/1}^{CLP} + A_{0/1}$]." (Floridi, 2011b, p. 197).

This much established a definition of the truthful nature of semantic information can be provided:

- i. 'the beer is in the fridge' qualifies as semantic information if and only if
- ii. 'the beer is in the fridge' is true; (ii) is the case if and only if
- iii. 'yes' is the correct answer to (i.e. correctly saturates by correctly verifying and validating) the question 'is the beer in the fridge?'; (iii) is the case if and only if
- iv. 'is the beer in the fridge?' + 'yes' generate an adequate model m of the relevant system s ; (iv) is the case if and only if
- v. m is a proxy of s and proximal access to m provides distal access to s ; and finally (v) is the case if and only if
- vi. reading/writing m enables one to read/write s ." (Floridi, 2001, pp. 198-199).

According to Floridi (2011b) the reduction of truth to correctness frees the account of an ontological commitment – i.e. a commitment to some sort of relation or correspondence to real existing entities. This gives the account an advantage over, for example, correspondence theory – which is dependent on such corresponding relations – and it enables the correctness account to deal with fictional truths.

To deal with fictional truths is just a matter of setting the CLP parameters in accordance with the fictitious universe one wishes to deal with. Thus, the fictitious universe itself is the level of abstraction (e.g. the universe of Sherlock Holmes), the context can be a specific episode or

¹⁰² The read/write terminology is from "computer science, where accessibility refers to the actual permission to *read* (technically, *sense* and *retrieve*) and/or *write* (again, technically *modify* and *record*) data as a *physical* process." (Floridi, 2011b, p. 197).

part of that universe (e.g. a specific story about Sherlock Holmes), and the purpose can be to simply discuss what is going on in that universe (e.g. to discuss whether Sherlock Holmes and Watson are friends), but the purpose could also be analyses of Sherlock Holmes' reasoning and deductions as cases for research in logic. Therefore, although Sherlock Holmes and Watson are not real living persons, it is still true, within the fictitious universe of Sherlock Holmes (the LoA), that they are friends (depending on how one defines friendship – which could be a purpose of its own). In this way, Floridi dispenses with the theories of meaning put forth by Frege and Russell – theories which both attach to the notion of truth in terms of truth-values. In contrast to Frege's (1892) account of sense and reference, a fictional name, on Floridi's account, does have a reference within the specific LoA, wherefore a truth-value can be ascribed. And in contrast to Russell's (1905) theory of definite descriptions, the ascribed truth-value does not have to be 'false'.

The correctness account's possibility of dealing with fictional truths along with other semantic content that does not have a direct correspondence to physical entities in the world is, according to Van der Vere Martens (2015), one of the attractions of Floridi's interpretation of philosophy of information to information studies (or more precisely, library and information science):

"Of particular interest to LIS is Floridi's "correctness" theory of truth, which provides for those special instances in which categories of fictional, empirical, ethical, modal, dispositional, metaphorical, and other semantic content are of interest (...). All these can be dealt with as *information*, provided that attention is given to the levels of abstraction at which they are being analyzed. For instance, "Dr. Watson is Sherlock Holmes's best friend" is true within a level of abstraction that deals specifically with that fictional world (...), but not within a level that deals with historical data from the nineteenth century." (Van der Vere Martens, 2015, p. 332).¹⁰³

The correctness theory of truth is developed to fit into Floridi's (2004a, 2011a) constructionist view of knowledge and information where "[w]e neither discover nor invent the world; we design it. So we understand it derivatively, only insofar as we understand its models. (...) This is neither a realist nor an antirealist but a *constructionist* view of information." (Floridi, 2004a, p. 663). The constructionist view with emphasis on 'designing' and 'engineering' forces Floridi to argue against a correspondence theory of truth in the first

¹⁰³ Bear in mind that, according to Floridi (2011b) attention given to the level of abstraction is not enough, it is an "*absolutely crucial fact* that the whole analysis must be conducted by paying careful attention to the LoA, the context and the purpose of the corresponding questions." (Floridi, 2011b, p. 204).

place. A view of truth as a correspondence between truth-bearers and ‘the world as it is’ does not fit with a view of the world as something which agents design. When the world is viewed as something which agents continuously design there is no ‘world as it is’, wherefore, there is nothing which the truth-bearers can correspond to. Thus, it is not just an advantage but an actual requirement for Floridi’s correctness theory of truth, that it does not have an ontological commitment.

Floridi puts emphasis on the correctness theory as a pragmatic theory of truth as it introduces an active agent who holds the information and interacts with it. Such pragmatic features are also present in the conclusions, which Fox (1983) draws from his analyses of ‘inform’ and ‘misinform’ (analyses conducted as part of the investigation of ‘information’ and ‘misinformation’; cf. section 3.4).

4.2.6 Default truth-values

According to Fox (1983) even though truth is not a semantic feature of ‘inform’ it is still reasonable to associate ‘inform’ with truth because of Grice’s Cooperative Principle and its maxims (cf. section 2.3):

“More specifically, if ‘inform’ were used in lieu of ‘misinform’ when P is believed to be false, then either one would not be as informative as required by maxim (1), or one would be obliged to state further that P is false, which violates maxim (4). Thus, in general, if the speaker believes P to be false, it is proper and conventional to use ‘misinform’. Otherwise (meaning those cases in which, to the best of the speaker’s knowledge, P is true), one uses ‘inform’. By default, then, ‘inform’ is generally associated with truth (or with information believed to be true). Thus there is a perfectly reasonable explanation for the traditional associations between informing and truth, and misinforming and falsehood even though the former association is not a semantic feature of ‘inform’.” (Fox, 1983, p. 160).¹⁰⁴

In addition to the conversational maxims the conclusion stems from the fact that on Fox’s account both ‘inform’ and ‘misinform’ entail ‘being in a position to know that P’. ‘Being in a position to know that P’ means that an agent only can inform or misinform if he is an epistemic authority with regard to p and have “adequate justification to support a belief concerning whether p is the case.” (Fox, 1983, p. 179). However, the agent need neither know

¹⁰⁴ Fox paraphrases Grice’s categories of maxims into a briefer list of four maxims: “1. Be as informative (but no more so) than is required for the current purposes of the exchange. 2. Be truthful. 3. Be relevant. 4. Be brief, clear, and orderly.” (Fox, 1983, p. 128).

nor believe p , he just have to have the right preconditions for knowing or believing whether p . Thus, if an agent consults a watch he is in a position to know what time it is, regardless of whether he gets the time right or not (e.g. regardless of whether he misreads the digits).

The default understanding of ‘inform’ as true might explain some of the reason why someone thinks that information requires truth – they are simply misled by the pragmatic features of ‘inform’ as well as what it is to ‘be in a position to know’. Although the scope is different, as there are no pragmatic features connected with natural information, Scarantino and Piccinini (2010) make a similar point when they argue that

“[t]he goal of an information-based account of knowledge has led many to neglect cases of probabilistic natural information that do not yield knowledge, yet fully qualify as information. In other words, the true thesis that information is *capable* of yielding knowledge gets confusedly mixed with the false thesis that information yields knowledge *in all cases*.” (Scarantino & Piccinini, 2010, p. 317).¹⁰⁵

Fox’s argument and explanation for why ‘inform’ par default is conceived as yielding truth although truth is not a feature of ‘inform’ might apply to ‘misinform’ (and ‘disinform’) and ‘falsity’ as well. On Fox’s (1983), Dretske’s (1981), and Floridi’s (2005a, 2011b) accounts misinformation is defined as necessarily false – i.e. as false semantic content. Thus, ‘misinform’ is also connected to falsity. In the explicanda of misinformation falsity is present as well, however, it is not a necessary condition and most of the explicanda of misinformation are defined in terms of misleadingness rather than falsity or at least as a supplement to falsity. On Fallis’ (2009, 2011, 2014, 2015) accounts of misinformation falsity is not a requirement either and misinformation is defined as inaccurate information or inaccurate and misleading information. Thus, on these accounts ‘to misinform’ would be to provide false and/or misleading information. With disinformation (and thereby ‘disinform’) it is different. All accounts of disinformation, both the explicanda and the philosophical accounts require falsity for disinformation – except the accounts developed by Fallis (2009, 2011, 2014, 2015) where disinformation can be true. Fallis’ notion of true disinformation is a legitimate derivation from the literature on lying, misleading, and deceiving based on Gricean implicatures, wherefore it must be accepted. If disinformation is a kind of misinformation – i.e. the kind which is intentionally misleading or deceptive – then misinformation cannot be strictly false, only inaccurate and misleading.

¹⁰⁵ Probabilistic natural information = “(PRT_N) If a signal s being F carries natural information about an object o being G , then $P(o \text{ is } G | s \text{ is } F) > P(o \text{ is } G | \sim(s \text{ is } F))$.” (Scarantino & Piccinini, 2010, p. 317).

However, when ‘misinform’, ‘disinform’, ‘misinformation’, and ‘disinformation’ are most commonly associated with falsity it might be due to a default ascription of truth-value based on the Gricean maxims – like the default ascription of truth to ‘inform’. Misinformation and disinformation are defined in terms of ‘misleading’ and ‘deceiving’ which in the literature of lying, misleading, and deceiving are defined as wrong, bad, immoral actions (Adler, 1997; Fallis, 2010; Mahon, 2008; Stokke, 2013; and Webber, 2013). Therefore, ‘to misinform’ and ‘to disinform’ must also be wrong, bad, immoral actions. Grice’s Cooperative Principle and its maxims describe the conditions under which verbal communication is possible, good, orderly, and successful. By every definition presented misinformation and disinformation violates the maxim of Relation (“Be relevant”), all the maxims of Manner (“Be perspicuous”, “Avoid obscurity of expression”, “Be brief”, “Be orderly”), and the two maxims of Quantity (“Make your contribution as informative as is required”, “Do not make your contribution more informative than is required”) simply by their misleadingness. Therefore, it is reasonable to think that misinformation and disinformation par definition also violates the maxims of Quality (“Try to make your contribution one that is true”, “Do not say what you believe to be false”, “Do not say that for which you lack adequate evidence”) and therefore are inherently false, although Fallis (2009, 2011, 2014, 2015) has shown that falsity is not a requirement for disinformation. Thus, due to the Cooperative Principle and its maxims ‘inform’ is by default associated with truth although truth is not necessary for ‘informing’ to take place (Fox, 1983), whereas ‘misinform’ and ‘disinform’ by default are associated with falsity although falsity is not necessary for ‘misinforming’ and ‘disinforming’ to take place.

4.3 Truth and intention

The notion of truth is significant in connection to information, misinformation, and disinformation within philosophy of information. The debates about especially information but also misinformation and disinformation often concern truth-values – does information require truth and do misinformation and disinformation require falsity? Or can some of the notions be alethically neutral? The notion of truth itself is rarely discussed and it is seldom specified what is meant by ‘truth’, when truth is a requirement for information. Floridi (2011b) seems to be the only exception when he develops his own theory of truth, the correctness theory. However, some sort of correspondence or deflationary approach seems to be presupposed by most of the theories of information, misinformation, and disinformation examined.

Whether information, misinformation, and disinformation come with specified truth-values or whether they can be alethically neutral partly determine how the three notions are and can be interconnected – i.e. whether they are different in kind (Budd, 2011; Dretske, 1981; and Floridi, 2005a, 2005b, 2011b) or whether they are varieties or kinds of one another (Fallis,

2009, 2011, 2014, 2015; Fetzer, 2004a; Fox, 1983; and Scarantino & Piccinini, 2010). Fallis' notion of true disinformation challenges the truth-false dichotomy for information vs. misinformation and disinformation – a dichotomy which is also challenged by Fox (1983) in his argumentation for the default association between 'informing' and 'truth'. In Fox's argumentation intentions become relevant for the definition of information because the intention to inform (in Fox's sense) about something true is present even though it might turn out to be false. The association between information and truth arises due to the expected intentions of the informer – both due to the Cooperative Principle and to the three intentions which Grice describes as necessary for nonnatural meaning to be present in communication. Thus, in order to further investigate the interconnections of information, misinformation, and disinformation 'intention' as the last asymmetry must be taken into account.

Ch. 5 Intention

5 Intention¹⁰⁶

When it comes to the notion of intention there are roughly two large bodies of literature which are relevant to the notions of information, misinformation, and disinformation – the philosophy of intention and the philosophy of action. In connection to these there is also the body of literature on intentionality.

Roughly, there seems to be two discussions when it comes to the notion of intention (when the discussions concerning action are not taken into account). The first discussion is concerned with the definition of intention and whether it should be defined in terms of belief, desire, judgment, or something else entirely. The second discussion is concerned with the relations between three guises of intention (Setiya, 2009; Wilson & Shpall, 2002):

- 1) Intention for the future;
- 2) The intention with which someone acts; and
- 3) Intentional action (intentionality)

as well as how to find unity in the three guises – e.g. whether they are to be explained in terms of one another or whether one of the guises is primary and in that case which one (Setiya, 2009).

¹⁰⁶ Some of the sections in this chapter have been presented in a shorter version in Søre (submitted) *Intention and Misleading*.

The relations between the three guises and how to find unity between them are not of concern here as it is a more fundamental discussion within Philosophy of Intention, Philosophy of Action, and on the border of Philosophy of Mind. As stated by Wilson and Shpall (2002) “the concept of ‘intention’ has various conceptual inflections whose connections to one another are not at all easy to delineate, and there have been many attempts to map the relations between intentions for the future, acting intentionally, and acting with a certain intention.” (Wilson & Shpall, 2002, introduction). However, the structure of the discussion resembles the overall purpose of this dissertation – to discuss the connections (or lack thereof) between information, misinformation, and disinformation and how (if possible) to shape a unified conceptualization of them, in philosophy of information¹⁰⁷. The purpose is to clarify how intention can be understood and defined in order to analyze and discuss what is entailed when misinformation and disinformation are defined in terms of intention or intentionality.

Thus, the exact definition of intention (a definition to which there is no consensus) is not of the utmost importance as the scope is neither to provide a new definition of intention nor to explore which notion of intention is present (if present at all) in the various definitions of information, misinformation, and disinformation. Rather, the scope is to analyze and discuss how ‘intention’ and ‘intentionality’ influence the notions of information, misinformation, and disinformation. That is, to investigate the connections between the three notions through the aspect of intention/intentionality as the third asymmetry. As part of these analyses and discussions investigations of the implications, which the asymmetry – in terms of intention/intentionality – has for the creation of a unified conceptualization of information, misinformation, and disinformation are conducted. Furthermore, the distinction between intention and intentionality is crucial in order to understand what happens when somebody misinforms as well as to capture all instances of disinformation (Fallis, 2014).

5.1 Theories of intention

The notion of intention is debated as much as the notions of meaning and truth and many different theories of intention and action have been put forth. According to Setiya (2009) the theories can be classified on two axes, where the first is how they find unity in the three guises of intention, and the second is how “they understand the relation between intention and evaluative thought” (Setiya, 2009, introduction). In contrast to ‘meaning’ and ‘truth’ there

¹⁰⁷ At the moment it seems that information ‘is treated as primary’ and not explained in terms of misinformation and disinformation, whereas misinformation and disinformation are explained in terms of information.

is nothing in the different theories of information, misinformation, and disinformation, which indicate a specific understanding of intention or intentionality – with Dretske (1981) and his orders of intentionality as the only exception. Thus, the purpose is not to examine which notion of intention or intentionality (if any) is present within the various accounts and definitions of information, misinformation, and disinformation. Rather the scope is to present some general understandings of what intention and intentionality can be in order to investigate which implications these two notions have for the definitions of information, misinformation, and disinformation and their interconnections.

5.1.1 Intentionality

Intentionality is best understood as directedness towards something, a “pointing towards or attending to some target” (Jacob, 2003, section 1) and it is closely linked to the intersection between philosophy of mind and philosophy of language (Jacob, 2003). According to Jacob (2003) intentionality is a feature of mental states such as desires, hopes, judgments, beliefs, intentions, etc. – i.e. intentionality is a feature of intention¹⁰⁸. In other words, the intentionality as a feature of intention is the directedness towards the intended action or state of affairs. However, intentionality can also be the awareness or foreseeability of consequences of actions, where the action is intended or desired but the consequence might or might not be intended or desired – “when *S* is doing *A* intentionally, *S* knows that she is doing *A*.” (Setiya, 2009, p. 10). Because the consequence, or side effect, is foreseeable the action is intentional even though it might not be intended. This also means that intentional action cannot be accidental – for an action to be intentional it must at least be foreseeable if not actually intended. Intentionality is also the directedness of an action even though the outcome of the action does not correspond with the initial intention, as when an agent misinforms with the intention of informing. The propositions uttered are uttered intentionally because the agent does not produce the sounds, words, and sentences by mistake, but the intention of informing is not met due to the inaccuracy or outright falsity of the intentionally uttered propositions.

As intentionality in the sense of directedness or foreseeability is a feature of mental states in general it follows that intentionality can be present without intention either as the intentionality of some other mental state or as an intentional action derived by foreseeable consequences or side effects, which themselves are not intended. This is the case with side effect disinformation as described by Fallis (2014) (cf. section 5.2).

¹⁰⁸ Whether intention is actually a mental state is up for debate (cf. Setiya, 2009).

In addition to intentionality in the directedness and foreseeability senses, there is also intentionality in the aboutness-sense (Adams, 2015, personal communication). Intentionality in the aboutness-sense is the kind of intentionality exhibited by the content of some signal, state, or event – for example, a cognitive structure – and it means that the content is about something. This is the kind of intentionality which Dretske (1981) describes in his attempt to naturalize intentionality – i.e. to show that non-mental commodities (i.e. information) can exhibit some sort of intentionality (the 1st order) and thereby to show how meaning can arise (by structures who exhibit the 2nd or 3rd order) (cf. section 3.2). Thus, intentionality in the aboutness-sense is connected to meaning in the sense, that although information has aboutness independent of ‘meaning’ and agents, agents cannot assess the aboutness of information without ‘meaning’ – i.e. agents need ‘meaning’ in order to determine what the information is information about. Naturally, intentionality in the aboutness-sense can, in the same way as intentionality in the directedness- or foreseeability-sense, be present both with and without an intention. Thus, information, misinformation, and disinformation all exhibit some degree of intentionality, i.e. intentionality in the aboutness-sense, because they have content – information, misinformation, and disinformation are about something.

5.1.2 Intention as belief

When it comes to the conception of intention there are many different approaches. Some are reductive and hold that intention can be explained in terms of mental states such as desire and belief. Others are causal where intentions are explained in terms of judgments or settledness on a certain desired action, which is believed to be obtainable. Proponents of the causal explanations (e.g. Mele, 1992; Davidson, 2001) criticize the reductive approaches – exemplified by the Cognitivist approaches (e.g. Setiya and Grice – cf. Wilson & Shpall, 2002) – because they believe that there is more to intentions than mere belief or desire; for example a settledness on performing the desired action (Mele, 1992) or an evaluative judgement (Davidson, 2001). Intentions can also be explained as goal oriented in terms of planning or they can be defined as reasons for acting. Still others just define intention as a propositional attitude (Jacob, 2003; Mele, 2010; Setiya, 2009; and Wilson & Shpall, 2002). For the purpose of intention in connection to disinformation, misinformation, and information the Cognitivist approach seems sufficient. The Cognitivist approach is a reductive approach that identifies intention as belief (strong interpretation), or defines the intention to do something as partly

constituted by the belief that one will actually do what one intends (weak interpretation)¹⁰⁹ (Wilson & Shpall, 2002).

‘Misinforming’ and ‘disinforming’ are not actions in a purely physical sense (as, for example, running). They might be oral acts or actions through writing¹¹⁰ – i.e. communication – where the action is bound in the propositional content separated from the medium with which it is delivered (speech or writing). What seems to be at play is the agent’s belief about the truth-value of the proposition or the desire to, and belief that, he will mislead. In connection to communication and intentions, Grice (whom Wilson and Shpall categorize as having a weak cognitivist approach) states that

“[i]t is general true that one cannot have intentions to achieve results which one sees no chance of achieving; and the success of intentions of the kind involved in communication requires those to whom communications or near communications are addressed to be capable in the circumstances of having certain thoughts and drawing certain conclusions” (Grice, 1989, p. 98).

Furthermore, Mele (2010) has stated that “[i]t is generally agreed that intentions are closely linked to desires – especially *action desires*, desires to do things – and beliefs.” (Mele, 2010, p. 108) – the disagreements are about how intentions and desires or beliefs are linked¹¹¹. Setiya (2009) puts it somewhat differently: “It is a matter of consensus in the philosophy of intention that intending to do *A* entails wanting to do *A*, in the motivational sense for which the ‘primitive sign of wanting is *trying to get*’ (...). Doubts about this entailment are attributed to ambiguities in ‘desire.’” (Setiya, 2009, section 3).

Therefore, the definition of intention for the present purpose, will be the Gricean weak Cognitivist approach, which in Mele’s (2010) words says that

INT_{GR} “having an intention to *A* requires believing that one (probably) will *A*” (Mele, 2010, p. 108).

¹⁰⁹ A proponent of the Strong Cognitivist approach is Setiya, whereas Grice is a proponent of the Weak Cognitivist approach (Wilson & Shpall, 2002).

¹¹⁰ Oral acts are physical in the sense that the agent has to move his lips and use his tongue in order to produce the sounds which constitute speech. Writing is physical in the sense that the agent has to hold the pencil and move his hand and arm, or press the keys. These physical acts of speaking and writing are different from the acts of communication – e.g. ‘misinforming’ or ‘disinforming’ – which are bound in the content uttered or written – i.e. the propositional content.

¹¹¹ Mele is a proponent of a causal explanation of intentional action, where intentions are partly belief partly settledness on a specific action which corresponds to the belief (Mele, 1992).

Following Grice (1971) more closely, it is the strict sense of ‘intends’ where ‘X intends to do A’ implies that X believes that he will A (Davidson, 2001). In other words, an agent cannot intend to buy a newspaper if he knows that he will in fact not buy it or if he believes that he will not buy it. He cannot intend, either, to buy the newspaper if he believes it to be impossible to buy a newspaper. If an agent intends to buy a newspaper it implies that he believes that he will buy it. If he believes that something might prevent him from buying the newspaper, he cannot intend to buy it, he can wish or hope or desire to buy it, etc.

5.2 Misleading – Intention and intentionality

The difference between intention and intentionality is especially important when it comes to misinformation – as well as disinformation defined as intentional misleading (Fallis, 2014). In the case of misinformation the misleading or falsity is unintended, which means that some other intention must have been present when the misinformation was disseminated. The intention present was most likely the intention to inform or provide information (i.e. true or accurate information) but as the ‘information’ is false or inaccurate it is also misinformation or misinformation instead – depending on whether one allows information to be false¹¹². However, the false or inaccurate ‘information’ (i.e. semantic content) is disseminated intentionally (in the directedness sense) as the utterance of the propositions in question is not accidental – when the utterance is intentional so is the misinforming. It is not the mere utterance which renders the propositions misinformation; it is the falsity or inaccuracy caused by the agent’s ignorance or fallibility which determines that it is an instance of misinformation and thereby a case of misinforming. This means that a significant feature of misinformation is, that the intention present (i.e. the agent’s belief as to what he is doing and thereby will achieve) does not match the intentional action, which he is actually doing.

In the case of disinformation as intentional misleading it is the same mechanism which allows the encapsulation of side effect disinformation (i.e. the mechanism that the intention present might not match the actual intentional misleading; cf. Fallis, 2014, 2015). The mechanism behind side effect disinformation bears some resemblance to the Principle of Double Effect in moral philosophy and ethics – the disinformation is an inevitable consequence and side effect of the initial action but it is a consequence which, although foreseeable, is not intended. The action has some other intended goal of which the disinformation is a side effect. For instance, if teachers create fake webpages with inaccurate and misleading content in order to teach their students how to critically examine what they find on the internet, then they do not

¹¹² Cf. section 5.3 as well as section 1.2 for a discussion of when the different notions can be applied.

intend actually to mislead anybody but it is a foreseeable side effect that someone might actually be misled. The foreseeability of the misleading renders the misleading intentional and the fake webpages are a case of side effect disinformation (Fallis, 2009, 2011, 2014, 2015). It is the foreseeability of the misleading which marks the difference between side effect disinformation and misinformation. The misleadingness of misinformation is never foreseeable for the agent who provides the misinformation even though the misleading is intentional.

As Fallis' (2014)¹¹³ application of 'intentional' (instead of intention) to misleading is meant to capture side effect disinformation, it follows that most often the misleading is intended as well as intentional. When the misleading is also intended it follows from the definition of intention (cf. section 5.1.2) that the agent believes that he probably will mislead. Thus, an agent cannot disinform (i.e. transmit disinformation to other agents) unknowingly because then he would not have the belief that he would succeed in misleading them. Even if the misleading is only intentional and not intended the agent cannot disinform unknowingly as intentionality is a directedness – a feature of mental states which requires awareness or foreseeability (Jacob, 2003).

In other words, if α is misled by the disinformation p (i.e. α now believes p) and α transmits p to β , then p is not disinformation anymore because p is neither intended nor intentional misleading. If α believes p , then it is not foreseeable for α , that p is misleading whereby it cannot be intentional misleading and it certainly is not intended either. The question is what p is when it reaches β ? Now, if the truth-value of p is false or p in some way is inaccurate, it becomes unintended misleading and therefore misinformation. However, if the truth-value of p is true – as in a case of true disinformation (Fallis, 2009, 2011, 2014, 2015), the question is not that easy to answer. For example, when a bank robber, Carl, is asked by the police if he knows where the bank robber is, and he answers 'not far away'¹¹⁴, then by use of a Gricean implicature he falsely implicates that he is not the bank robber, wherefore, the otherwise true statement, that the bank robber is not far away, becomes misleading. If the police disseminate this piece of disinformation to their colleagues or others, what does it become then? Does it become information? Or should it be labelled true misinformation due to the continuance of

¹¹³ In Fallis (2009, 2011, and 2015) the distinction between intention and intentionality is alluded to, through the use of 'foreseeability'. However, the explicit use of 'intentionality' in the definition of disinformation appears in Fallis (2014). Cf. section 1.5.2, note 37 for a note on the chronology of the articles.

¹¹⁴ The misleading in this instance of true disinformation works due to the false implicature of 'not far away' which implicates that Carl is not the bank robber. The structure of the example is based on the example given in Fallis (2014).

the misleadingness, thus introducing a new variety of misinformation? And is such a new variety of misinformation – i.e. true misinformation – even possible?

On the accounts of misinformation provided by Dretske (1981), Floridi (2005a, 2005b, 2011b), and Fox (1983), where misinformation is defined by falsity, true misinformation is a contradiction in terms. However, in regard to Dretske's and Floridi's accounts of information, there would be something odd about saying that true disinformation can become information when the intention to mislead is gone, especially when information is defined as an objective commodity independent of agents (Dretske, 1981) or at least independent of informees (Floridi, 2005a). The oddness occurs because even though the intention to mislead is gone, and it therefore cannot be disinformation, it is still misleading, and therefore cannot be information¹¹⁵ – in the example with the bank robber Carl, the false implicature, that Carl is not the bank robber, derived from the utterance 'not far away', is still present and therefore if police officer Allan is misled by Carl and tells his colleague Ben that Carl says, that the bank robber is not far away, Ben is likely to be misled as well, even though Allan has no intention to mislead and does not do so intentionally.

So in a case of true disinformation where the intention to mislead is not preserved through dissemination and the misleading is not intentional either, Ben is left with something true but misleading, which is not disinformation and cannot be misinformation either, if misinformation is defined by falsity. What Ben is left with seems to be true misleading information – though true misleading information is as contradictory as true misinformation, on the accounts where information is defined as an objective true entity¹¹⁶ and misinformation is defined by falsity. Such an entity cannot in itself be misleading as it just *is* – it exists in the world independent of agents (Dretske, 1981, 1983, 2008). The misleadingness therefore cannot be part of information itself. Information, as defined by Floridi (2005a, 2005b, 2007, 2011b) and Dretske (1981, 1983, 2008), might become misleading for an agent dependent on the agent's abilities to extract or understand the information (i.e. 'misleadingness' is a pragmatic feature, a defective relation between information and agent), but this is not what happens in the case of the provision of true misleading information described above. The original misleadingness is caused intentionally by an agent (Carl) through the use of something true (that the bank robber is not far away) in such a way, that he believes that he will succeed in misleading (the use of the false implicature).

¹¹⁵ On Fox's (1983) account of information it is perfectly legitimate to speak of 'misleading information'.

¹¹⁶ Dretske himself uses the term 'commodity' (Dretske, 1981, 1983, 2008).

In Fallis (2015) misinformation is defined as unintended inaccuracy instead of falsity. This makes a difference because true misinformation might then be possible¹¹⁷ – that something is inaccurate does not necessarily mean that it is false. For instance, if an agent is presented with a true subset of a larger picture but need the full picture in order not to be misled, then the true subset is inaccurate in relation to the larger picture. For example, if four clinical trials show that some drug has one effect on humans, whereas 50 other clinical trials show that the same drug has another effect on humans, then it is true but inaccurate and misleading if only the results of the first four clinical trials are referred to. If the lack of reference to the 50 other clinical trials is due to unawareness that they exist, then it is a case of unintended misleading by true inaccurate ‘information’. Thus, based on his notion of true disinformation, Fallis’ (2015) account of misinformation as unintended inaccuracy can be extended to incorporate true misinformation. According to this extension, Ben, in the example above, is left with true misinformation rather than true misleading information.

If it is Fox’s (1983) account of information and misinformation that is endorsed there can be no such thing as true misinformation because misinformation requires falsity in order to be misinformation. Instead, there is no inconsistency in saying ‘misleading information’ (Fox, 1983, p. 86+95) as information does not require truth. Recall that, although Fallis (2009, 2011, 2014, 2015) has showed – based on the literature on lying, misleading, and deceiving – that something true can be misleading, there is still a default association between misleadingness and falsity due to Grice’s Cooperative Principle and the maxims (cf. section 4.2.6)¹¹⁸. Furthermore, in the cases where what is literally said is true, the misleadingness arises due to the false implicature of what is said, wherefore misleadingness is connected to falsity in some way – either to the falsity of the semantic/representational content or to the falsity of the implicatum. On Fox’s (1983) account misinformation is a kind of information and informing and misinforming are both kinds of telling. In order to either inform or misinform the agent must be in a position to know that *P* (which does not mean that he knows that *P*) and whether *p* is labeled information or misinformation is dependent on the known or unknown truth-value of *p* (Fox, 1983). If *p* is known to be false it can be called misinformation, if the truth-value is unknown or *p* is known to be true, then it can be called information. This means that the term misinformation does not have to be applied when a proposition is false, but if the term misinformation is applied then the proposition has to be false.

¹¹⁷ Fallis (2015) does not address the possibility of true misinformation.

¹¹⁸ Fox (1983) does not address the association between misleadingness and falsity. However his examples and his argumentation that information can be misleading because it does not require truth suggest that misleadingness is roughly equivalent to falsity on Fox’s account.

As for the case with Carl, Allan, and Ben – where the original true disinformation misleads Allan – Allan, who is unaware of p 's misleading features, disseminates p to Ben who is now left with something true and misleading. There are two strategies which can be taken in order to resolve the puzzle; either misinformation must be defined as unintended inaccuracy which leaves room for true misinformation, or it must be possible to speak of misleading information in order to capture the cases where some content is true and misleading. The possibility of misleading information requires that information can have false implicata and still be information. However, as argued in section 3.5.1, the implicatum counts as part of the representational content – it is part of the nonnatural meaning, the speaker-meaning – therefore, a definition of information where information can be misleading must be able to capture falsity. The simplest way is to follow Fox (1983) and Fallis (2009, 2011, 2014, 2015) and define information as alethically neutral.

5.3 Belief and interconnections

When intention is defined in terms of belief and intentionality is defined in terms of directedness and foreseeability, which also entail a belief, it follows that disinformation (defined as intentional misleadingness) and misinformation (defined as unintended misleadingness by inaccuracy) are linked to beliefs. In order to disinform the agent must at least foresee that he is likely to mislead – often he intends to mislead as well – and in order to foresee the misleading or intend it the agent must believe that he probably will mislead (that he is likely to mislead). In order to misinform, on the other hand, the agent must un-intend to mislead and in order to un-intend the agent must have a non-belief that he will mislead. The agent who misinforms most likely has the belief that he will actually inform¹¹⁹ the receiver and thereby has the intention to do so.

5.3.1 Veridical information and interconnections

Information, on the other hand, in its semantic and veridical interpretation (Dretske, 1981, 2008; Floridi, 2005a, 2005b, 2007, 2011b) is not linked to beliefs in this way. Belief does not come sneaking in the backdoor disguised as intention or intentionality (in the directedness- or foreseeability-sense) in the definition of information. In Dretske's and Floridi's¹²⁰ accounts of information it is intentionality in the aboutness-sense which is present. Intentionality in the

¹¹⁹ I.e. 'inform' in a default understanding as truth-providing (cf. section 4.2.6).

¹²⁰ Floridi (2005a, 2005b, 2007, 2011b) does not address intentionality in the aboutness-sense. However, his notion of information is a semantic notion, it is about something, therefore, it must have intentionality in the aboutness-sense.

aboutness-sense does not entail belief as, at least on Dretske's (1981) account, it is present within the information carried by a signal independent of agents. Beliefs first enter the picture afterwards as cognitive structures which discriminate between the pieces of information carried by the signal. Thus, belief as a propositional attitude is capable of rendering the information 'that P' into knowledge (Dretske, 1981). Although in different terms, Hintikka (2007) points to the same idea about information as objective content which can be picked up and discriminated by different cognitive structures or propositional attitudes:

"[W]hat the content of a propositional attitude amounts to can be thought of as a certain item of information. In attributing different attitudes to agents, different things are said about the information – for instance, that it is known, believed, remembered, and so on. This fits in well with the fact that the same content can be known by one person, believed by another, remembered by a third one, and so on. This idea that one and the same objective content may be the target of different people's different attitudes is part of what Frege (...) was highlighting by his notion of *the thought*." (Hintikka, 2007, p. 17).

Hintikka (2007) does not have a truth requirement for information and Frege's (1892) notion of the thought is roughly equivalent to a proposition (in an interpretation of proposition as objective entities) and not to Dretske's notion of information. However, the idea of information as an objective and agent-independent entity or commodity which can be picked up by different cognitive structures is more or less the same.

Thus, if a Dretske-Floridian notion of information as veridical and objective is endorsed the asymmetry in terms of intention and intentionality between information, misinformation, and disinformation will affect the interconnections between, and shape the conceptualization of, information, misinformation, and disinformation in a specific way.

In the theories of information put forth by Dretske (1981, 1983, 2008) and Floridi (2005a, 2005b, 2007, 2011b) misinformation is not a kind of information it is something different – it is that which is false as opposed to information as that which is true – despite whether the two are direct opposites (i.e. Floridi's accounts) or whether information is prior to misinformation (Dretske's account). Though, due to the asymmetry in terms of intention and intentionality and the asymmetry in terms of truth as well as the realization that misinformation can be true, is agent-dependent, and entail some sort of belief, the distinction between information as that which is true and misinformation as that which is false cannot be upheld. The same holds for the distinction between information as true and disinformation as intended falsity as disinformation can be true (Fallis, 2014) and is agent-dependent.

5.3.2 Information as non-misleading

However, in the alethically neutral definitions of information as communication (both verbal, gestural, and by pictures) (Fallis, 2009, 2011, 2014, 2015; Fox, 1983¹²¹; and Scarantino & Piccinini, 2010) intentions and intentionality (both in the directedness-/foreseeability-sense as well as the aboutness-sense) are present. Thus, if this conception of information as alethically neutral and communicative in nature is endorsed, the asymmetry in terms of intention and intentionality and the asymmetry in terms of truth will affect the interconnections differently and shape a different conceptualization of information, misinformation, and disinformation (cf. section 6.5 for the full discussion). However, as information as alethically neutral are connected to notions of misinformation and disinformation that are also alethically neutral there must be something else which can be used to distinguish between the three notions – i.e. some feature which is necessary in order to fully explore the interconnections between information, misinformation, and disinformation, when they are all defined as alethically neutral.

Information – as well as misinformation and disinformation – plays an important part in communication. When agents communicate intentions are present or the communication is at least an intentional action. These intentions or the intentionality in the directedness- and foreseeability-sense are built into the definitions of misinformation and disinformation in terms of unintended misleadingness and intentional misleadingness. That is, if an utterance¹²² is misleading the intentional or un-intentional character of the misleading is specified. However, if an utterance is non-misleading (e.g. information – insofar as information is actually non-misleading) the intentional or un-intentional non-misleading is not explicit in the definition of information. None of the definitions of information examined require explicitly that information is non-misleading although non-misleadingness must be a feature of information: both if the Dretske-Floridian view is accepted, as something which is true and objective cannot be misleading in itself as it just exists in the world; but also if information as alethically neutral is accepted as, there must be some feature which distinguishes information from misinformation and disinformation. When misinformation is defined as unintended misleading and disinformation is defined as intentional misleading, true misinformation and

¹²¹ Fox (1983) explicitly states that telling does not require the intention to tell on part of the teller as slips of the tongue can be acts of telling as long as they are heard and understood by the one being told. Fox does not explicate whether informing, as a kind of telling, does not require intention either. However, it seems that since informing that *P* requires an epistemic authority who is in a position to know whether *p*, informing is at least an intentional action in the directedness-sense.

¹²² It can be anything with representational content; an utterance, a tweet, a pHEME, a picture, etc.

true disinformation are possible (cf. section 5.2). Thus, truth in itself does not prevent misleadingness even though it is exactly this requirement – in Dretske’s and Floridi’s definitions of information – which secures knowledge insofar as the information is believed by an agent (Dretske, 1981, 1983, 2008; Floridi, 2011b). The non-misleadingness of information therefore must be an explicit requirement in order to secure the quality of the information.

According to Grice’s Cooperative Principle

“Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.” (Grice, 1989, p. 26)

and its maxims (cf. section 2.3.1), provision of information, as something truthful and unambiguous, is strived for in all sincere communication. However, communication can be ambiguous especially in – but not limited to – online social environments where the same meme, tweet, etc. can be present in different kinds of contexts at the same time detached from their spatio-temporal origin. Sometimes, as with the phemes in the PHEME-project, the misleadingness or non-misleadingness, the truth or the falsity, of a single pheme differ from context to context whereby the same pheme can be information in one context but misinformation at the same time in another context. In order to determine which category the pheme must be mapped into (speculation, controversy, misinformation, and disinformation) (cf. PHEME, 2014) the spatio-temporal dimensions of the pheme as well as the intentions or the intentionality with which the pheme is created are relevant. Thus, intentionality in the directedness- and foreseeability-sense (as an addition to intentionality in the aboutness-sense) has emerged as a requirement for the definition of information. Hence two requirements, as found in the analyses of information in relation to misinformation and disinformation, must be incorporated into the definition of information; non-misleadingness and intentions/intentionality.

5.4 Intention, truth, and transitivity

The connections or relations between the aspects of ‘truth’ and ‘intention’ or ‘intentionality’ differ among the notions of misinformation and disinformation. In the definition of disinformation (Fallis, 2009, 2011, 2014, 2015) the intention or intentionality is not connected to the truth-value of the proposition – it is not intentionally false or intentionally true – it is intentionally misleading. With misinformation, though, the intention is most often connected with both the misleadingness and the truth-value in the way that the prefix ‘mis’ is added to ‘information’ because the intended non-misleadingness and intended truth-value do not match with the actual misleadingness and the actual truth-value (the only exception is

true misinformation where the truth-values match, but the intended non-misleadingness and the actual misleadingness do not match)¹²³. The prefix ‘mis’ is added retrospectively or by the agent who receives the misinformation if he knows the content to be false or inaccurate. An agent who disseminates misinformation will never know that this is what he does when he does it, because if he knows the content to be false and disseminates it anyway then he can no longer be said to be doing so unwittingly or unintended, and it would become disinformation (cf. section 1.2).

5.4.1 Disinformation as transitive

Intention/intentionality is what differ most significantly between the definition of disinformation as representational content which is intentionally misleading (Fallis, 2015) and the definition of misinformation as representational content that is inaccurate and misleading (Fallis, 2015). It is the intention or the intentionality of the action which determine whether it is an act of disinforming or misinforming, and thereby determines whether what is provided is disinformation or misinformation. In the example with Allan, Ben, and Carl it was seen that when some piece of disinformation (e.g. ‘He is not far away’) is passed on without an intention to mislead then it becomes misinformation instead. Thus, after the first link it is not disinformation anymore and it should not be possible to speak of the spread of disinformation, where the mechanism is often to deceive a few who will then pass on the deceptive content to others who might get deceived as well.

However, it does not seem fruitful not to be able to speak of the spread of disinformation, especially not in connection to online networks. People get misled by all sorts of things on the internet and although the ones who pass on the misleading representational content might not intend to mislead or deceive anybody, the ones who generated or created the misleading content in the first place might have done so with the intention to mislead and deceive. Thus, in order to speak of the spread of disinformation the original intention to mislead (or the original action with a foreseeable potential to mislead) must be preserved through transmission in some way or other. The preservation can be secured through transitivity. Transitivity is a principle from logic which states that if one can get from w to w' and from w' to w'' , then one can also get from w to w'' through w' ¹²⁴. Therefore, if disinformation is

¹²³ It is here presupposed that when someone misinforms, i.e. provides inaccurate and misleading representational content to someone else, what he actually strives for is to provide true and non-misleading representational content to that someone. This presupposition is derived from the specification in the definition of misinformation that it is unintended that the content is inaccurate and misleading.

¹²⁴ Transitive accessibility relation R : $(w - w') \wedge (w' - w'') \rightarrow (w - w'')$.

transitive then it can be preserved through transmission although the initial intention breaks down. In this way disinformation can spread through a network although the ones who pass on the disinformation (the different nodes in the network) do not share the initial intention to mislead and deceive. Thus, if A disinforms B and C that p and B and C believe p and transmit p to D, E, F, and G, then A disinforms D, E, F, and G through B and C, whereas B and C just misinform D, E, F, and G. The relations between A and B and A and C, respectively, are direct and immediate and the intention is to disinform B and C. In contrast, the relations between A and D, E, F, and G, respectively, are indirect and mediated and the intention is to disinform whoever receives the message through B and C.¹²⁵ In the Example with Carl, Allan, and Ben the transitivity of disinformation means that Carl directly disinforms Allan, Allan misinforms Ben, and Carl indirectly disinforms Ben through Allan. Thus, Ben ends up as disinformed although Allan only misinformed him. In this way, although the direct intention to mislead in principle breaks down through transmission, the indirect intention to deceive whoever receives the deceptive message can be preserved through transitivity as disinformation is transitive. Thus, disinformation can be preserved in transmission and the spread of disinformation is possible.

The transitivity of disinformation and the possibility of the spread of disinformation are what pose a threat to online social networks. That the direct intention to mislead breaks down and the further transmission of some misleading message becomes direct and immediate acts of misinforming and provision of misinformation, where the indirect intention to deceive anybody is preserved and the transmission therefore is indirect and mediated acts of disinforming and provision of disinformation makes it particularly difficult to distinguish between misinformation and disinformation in online social networks. Thus, if the original source is unknown, it might not be possible to determine, much less automatically detect, whether some short semantic structure (e.g. a message, a tweet, a Facebook update)¹²⁶ is misinformation, disinformation, or information, especially if what is detected is just truth and falsity (cf. section 5.2).

However, the transitivity of disinformation also poses a threat to automatic detection even when it is not truth and falsity that are detected. In Kumar and Geethakumari's (2014) algorithm-project it is not misinformation and disinformation (both defined in terms of falsity,

¹²⁵ Thank you to Emmanuel Genot and Staffan Angere for pointing out the need for transitivity of intentions in order to be able to speak of the spread of disinformation.

¹²⁶ Although a tweet and a Facebook update have a known tweeter and holder of the Facebook profile, it might be unknown where they got the message (the content) from in the first place (for example, if they picked it up outside the internet, the trail cannot be automatically detected within the social network structure).

inaccuracy, and deliberate misleading; cf. Introduction) that are detected, although the project is to detect and flag misinformation (as intended misleading). Instead, the proposed algorithm detects and measures trust and credibility of the original tweeter. The idea is that high trust and credibility, on behalf of the tweeter, lower the possibility of the tweet being misinformation (defined as synonymous to disinformation), whereas low trust and credibility, on behalf of the tweeter, raise the possibility of the tweet being misinformation/disinformation. Trust and credibility are measured on basis of the amount of retweets a given tweet generates. Here the idea is that, the more followers who systematically retweet tweets from a given tweeter indicate the level of trust and acceptance of that original tweeter. Thus, more retweets from many followers indicate high trust, credibility, and general acceptance, less retweets from few followers indicate low trust, credibility, and no general acceptance (Kumar & Geethakumari, 2014).

When it is trust and credibility of the original tweeter which are measured (based on numbers of retweets) – and not the content of the tweet itself which is analyzed – the ‘system’ (and hence the algorithm) gets vulnerable to disinformation. That is, if those who want to disinform can succeed in disinforming the original tweeters who have high credibility, then they can disinform through them due to the transitivity of disinformation. Thereby, the initial disinformers can bypass the algorithm, which will not be able to detect the disinformation. The ‘system’ (and hence the algorithm) also gets vulnerable to misinformation in Fallis’ (2014) sense (i.e. unintended inaccurate and misleading representational content). No real life agent is infallible wherefore he, once in a while, is likely to spread something which is inaccurate in some way or other (e.g. through performance error) and thereby to generate implicatures ‘by accident’ or, if one follows Bach (2006), to generate putative implicatures (cf. section 2.3.3).

5.5 Intention and unified conceptualization

Intention and intentionality are vital to the notions of information, misinformation, and disinformation and the distinctions between, and interconnections of, these notions. For example, intention/intentionality is crucial in the distinction between misinformation and disinformation as it is the un-intention to mislead which distinguishes misinformation from disinformation as intentional misleading when they both are defined in terms of misleadingness. That misinformation and disinformation are defined in terms of misleadingness rather than falsity is due to Fallis’ notion of true disinformation and the literature on lying, misleading, and deceiving, which enable the development of a true variety of misinformation as well. The entrance of true misinformation alongside true disinformation turns a truth-requirement for information insufficient for the distinction between information and mis-/disinformation. Instead misleadingness and non-misleadingness as features of

representational content which includes the implicata of Gricean implicatures seem to be the distinguishing features of mis-/disinformation and information. Therefore, non-misleadingness, both intentional and unintended, is proposed as a requirement for information. These requirements (i.e. non-misleadingness and intentionality in the directedness-/foreseeability-sense) challenge the present definitions of information and will change how the notions of information, misinformation, and disinformation are, and can be, interconnected as well as which unified conceptualization is possible. Thus, the final investigations of the interconnections of information, misinformation, and disinformation will result in the development of a unified conceptualization of the three notions.

Ch. 6 Unified Conceptualization

6 Discussions

In the previous chapters the conceptualizations of information, misinformation, and disinformation as developed by selected philosophers and information studies scholars within philosophy of information are presented, examined, and discussed. The main objective of these discussions is the interconnections between the three notions as Carnapian explicata with the fruitfulness for automatic detection as context. The conceptualizations are discussed in relation to one another within a Gricean framework of meaning and cooperation through the three asymmetries which concern aspects of 'meaning', 'truth', and 'intention'. The discussions aim at the development of a unified conceptualization of information, misinformation, and disinformation – a conceptualization where the interconnections of the three notions are taken into account and determine how the notions can be defined in accordance to one another such that the asymmetries are dissolved or are made coherent.

A few discussions remain before a unified conceptualization is developed. Therefore, I will discuss four central points which I have touched upon previously:

- Natural vs. nonnatural meaning
- Semantics vs. pragmatics
- Information and truth
- Disinformation and deception

These four discussions provide further insights and qualification for the conceptualization of information, misinformation, and disinformation.

6.1 The information booth

On Dretske's (1981, 1983, 1988, 2008) account of semantic information, information is defined as natural meaning in Grice's sense, it is an indicator-relation between a signal and a source and therefore true per definition and exhibits intentionality in the aboutness-sense. The signal contains information about its source. Systems and cognitive structures can have this information as their content and thus exhibit the 1st, the 2nd, or the 3rd order of intentionality depending on whether the structure can discriminate between the pieces of information carried by the source or not. 'Meaning' in Grice's nonnatural sense (although strictly semantic) is ascribed to the informational content by cognitive structures when they discriminate between pieces of information and select a discrete piece of information as their content. When cognitive structures (e.g. beliefs) have acquired some piece of information as their content this information can be communicated to other agents who can form beliefs on the basis of this information – i.e. they can generate cognitive structures which have this information as their content. In this way information on Dretske's account can be shared between agents but it only continues to be information as long as the informational link is preserved. The informational link is preserved as long as a given token, for instance the belief that *s* is *F*, is triggered by a signal which carries the information that *s* is *F*. If the token is triggered by another signal which does not carry the information that *s* is *F*, then it is a case of misrepresentation and misinformation can occur (cf. Ch. 3).

Dretske's (1981) emphasis on information as truthful is part of his project to develop an informational account of knowledge. When information is true per definition it can fulfil the truth-condition and replace the problematic justification-condition in the standard definition of knowledge, i.e. Justified True Belief (Dretske, 1983). Information as an objective true commodity, i.e. signals about actual states of affairs in the world is in itself justified, wherefore only belief has to be added in order to obtain knowledge. Thus, information believed by an agent (i.e. tokens where the informational link is preserved) amounts to knowledge for that agent. Dretske (2008) explains the truth-requirement and its relation to knowledge through an information booth-example:

“As the name suggests, information booths are supposed to dispense information. The ones in airports and train stations are supposed to provide answers to questions about when planes and trains arrive and depart. But not just *any* answers. *True* answers. They are not there to entertain patrons with meaningful sentences on the general topic of trains, planes, and time. Meaning is fine. You can't have truth without it. False statements, though, are as meaningful as true statements. They are not, however, what information booths have the function of providing. Their purpose is to dispense truths, and that is because information, unlike meaning, has to be true. If

nothing you are told about the trains is true, you haven't been given information about the trains. At best, you have been given misinformation, and misinformation is not a kind of information anymore than decoy ducks are a kind of duck. If nothing you are told is true, you may leave an information booth with a lot of false beliefs, but you won't leave with knowledge. You won't leave with knowledge because you haven't been given what you need to know: information." (Dretske, 2008, pp. 29-30).

However, the information booth does not provide information in the sense of natural meaning. When the information booth announces that the plane for London leaves at 5 pm this does not naturally mean that the plane for London leaves at 5 pm. It is not the case that if the information booth states that the plane leaves for London at 5 pm, then the plane cannot but leave for London at 5 pm. The times of departure for planes and trains in information booths mirror the scheduled times of departure not necessarily the actual times of departure. The times of departure in the information booth are not nomically or lawfully connected to the actual times of departure of planes the way signals normally have to be connected to events in order to carry information:

"When there is a lawful regularity between two events, statistical or otherwise, as there is between your dialing my number and my phone's ringing, then we can speak of one event's carrying information about the other. And, of course, this is the way we *do* speak. The ring *tells me* (informs me) that someone is calling my number, just as fingerprints carry information about the identity of the person who handled the gun, tracks in the snow about the animals in the woods, the honeybee's dance about the location of nectar, and light from a distant star about the chemical constitution of that body. Such events are pregnant with information, because they depend, in some lawfully regular way, on the conditions about which they are said to carry information." (Dretske, 1983, p. 56).

Thus, the information booth per se does not carry information about the actual times of departure of planes and trains. The semantic content in the information booth is dependent on nonnatural meaning. It is dependent on nonnatural meaning in the same way as traffic lights. That is, it is dependent upon conventions and intentions and does not reflect information-about-a-source in a lawful way. Recall that on Grice's account of meaning traffic lights are an instance of nonnatural meaning as there is some sort of connection to someone who means_{NN} something by these traffic lights, which is based on convention (cf. section 2.2 and Grice, 1957).

The convention comes in different varieties: in the traffic red means_{NN} stop; on electrical devices it means_{NN} off; etc. When the traffic light is red it carries information about the source, in Dretske's sense, that now it is the red light which is on, but it does not in this sense carry

information about the convention that red means_{NN} stop. This meaning_{NN} is not developed upon information in the way Dretske (1981) describes learning because there is no information in the first place that tells that red means_{NN} stop, this is something which humans have decided – i.e. there is no informational link between the convention and the red light. When the light is red it does not tell anything about whether there is other cars, or bikes, or people which need to cross the street it just tells the ones it faces (cars, bikes, pedestrians) that they have to stop no matter what. The red light does not carry information about the traffic. The shift to red is information about the source when it comes to the timing-function incorporated into the control of the lights, but this is not the kind of information needed in order to understand the meaning_{NN} of the red light.

It is not clear how conventions are treated on Dretske's account of information – can conventions, for instance, be events or states of affairs about which signals can carry information? When information is defined as an objective commodity independent of agents and when events and states of affairs, on Dretske's account, seem to be purely physical occurrences in the world, conventions cannot be states of affairs about which signals carry information. The difference between natural and nonnatural meaning is explained by Scarantino and Piccinini (2010) in terms of natural and nonnatural information: "Spots carry natural information about measles by virtue of a reliable physical correlation between measles and spots. By contrast the three rings on the bell of the bus carry nonnatural information about the bus being full by virtue of a convention." (Scarantino & Piccinini, 2010, p. 314).

The convention behind an information booth which provides times of departure and/or arrival for planes, busses, and trains is that it reflects the scheduled times of departure/arrival – i.e. it reflects the timetable – not necessarily the actual times of departure/arrival as delays can occur. The probability that the plane for London leaves at 5 pm is raised when it is stated in the information booth as the statement reflects the scheduled time of departure for the plane to London, however, the probability is not raised to 1 as the actual time of departure and the scheduled time of departure are not nomically related, i.e. the probability is not raised due to a reliable physical correlation. The probability is raised by virtue of the convention that planes, trains, and busses shall depart and arrive according to the schedule. On Dretske's account of information "[m]y gas gauge carries the information that I still have some gas left, if and only if the conditional probability of my having some gas left, given the reading of the gauge, is 1." (Dretske, 1983, p. 57). Scarantino and Piccinini's (2010) notion of natural information is a reply to and critique of Dretske's requirement that the probability that *s* is *F* is raised to 1 in order for a signal to carry the information that *s* is *F*. On Scarantino and Piccinini's account the probability is not necessarily raised to 1 but to some sufficiently high degree less than 1. Thus, the 'dispute' between Dretske and Scarantino and Piccinini about the

truth-requirement for information is actually a dispute about whether signals ever or often enough raise the probability of the occurrence of events to 1. Dretske's account may not satisfy Carnap's (1950) requirement of fruitfulness to a sufficient degree as it can be questioned with Scarantino and Piccinini in hand whether the majority of signals raise the probability of events to 1 (even when background knowledge is taken into account).

The distinction between natural and nonnatural meaning is important. It marks a difference between

- physical occurrences in the world that are independent of agents, and
- conventions and other agent-dependent constructs such as language.

I agree with Scarantino and Piccinini (2010) that natural and nonnatural meaning as the foundation for two different notions of information – natural and nonnatural information – marks an important distinction between two different ways of conceiving of information. These two conceptions are not mutually exclusive and they are both fruitful albeit for different purposes (cf. section 6.5.2).

6.2 Semantics vs. pragmatics

Floridi (2004b, 2005a, 2005b, 2007, 2011b) develops an account of semantic information based on alethic values (i.e. information is inherently truthful) rather than probability distributions. According to Floridi (2005b), one reason to prefer alethic values over probability distributions is the Bar-Hillel-Carnap Paradox (cf. section 4.2.5). Thus, Floridi (2004b, 2005a, 2007) calls his theory 'a theory of strongly semantic information' as a reply to Carnap and Bar-Hillel's (1952) 'theory of semantic information', which Floridi dubs 'weakly semantic information'. The 'weakly'-predicate refers to the alethic neutrality of Carnap and Bar-Hillel's account, whereas the 'strongly'-predicate refers to Floridi's requirement for truthfulness.

Floridi's (2004b, 2005a, 2005b, 2007, 2011b) theory of information is a strictly semantic account where 'meaningfulness' is a semantic feature of the content (i.e. sentence meaning) and not a pragmatic feature of the content (i.e. speaker-meaning). Dretske's (1981, 1983, 2008) account of information is also a strictly semantic account, where word- and sentence-meaning (referred to as nonnatural meaning) is not part of the information itself, it is ascribed to information by cognitive structures who pick up certain pieces of information as their discrete content (i.e. agents who form beliefs, etc.). The semantic aspect of Dretske's account is that it is an account of informational content – as opposed to Shannon's account which only deals with the amount of information – regardless of whether or not agents are present to ascribe meaning to this content. Fox's (1983) account of information (and misinformation) is different in the sense that it concerns communication. Fox's definitions of 'inform' and

'information' are cast in terms of telling and propositions. Fox's theory is about communicative acts where information is the content up for transmission. In his specification of what telling amounts to, Fox (1983) also specifies what does not count as telling:

"In metaphorical terms, shoes, rocks, bottles, and buildings, in fact virtually any sort of thing, can "tell" us that something is the case. For example, the mud on my shoes may "tell" you that I have been walking in the park. But literally speaking, my shoes (and other such items) don't talk, and they don't really tell anyone anything. Rather, we are able to draw conclusions about the way the world is based on the characteristics or behavior of things like shoes, rocks, bottles, buildings, and people." (Fox, 1983, p. 112).

What Fox dismisses as genuine cases of telling (and thereby dismisses as cases of informing) – wherefore they do not contain information – are exactly those cases that Dretske (1981) regards as pure instances of information (cf. section 6.1 – the quote from Dretske, 1983, p. 56). What counts as information for Dretske and Floridi does not count as information for Fox. However, the communicative aspect of Fox's theory and its consequences for what counts and what does not count as information is never addressed within philosophy of information. Fox's theory is never examined on its own terms, i.e. in the context of communication. Instead, Floridi (2005a, 2005b, 2011b) and The Π Research Network simply refer to Fox (1983) as one of the defenders of alethic neutrality for information. Floridi (2005a, 2011b) explicitly argues against Fox's (1983) theory of information and misinformation and lists Fox's conclusions about the relations between informing, misinforming, information, and misinformation as the ninth bad reason for holding that 'false information' is genuine information:

"FI.9) "'x misinforms y that p ' entails that $\neg p$ but 'x informs y that p ' does not entail that p [and since]... we may be expected to be justified in extending many of our conclusions about 'inform' to conclusions about 'information' [it follows that]... informing does not require truth, and information need not be true; but misinforming requires falsehood, and misinformation must be false." (Fox [1983], 160-1, 189, 193).

Objection: the principle of "exportation" (from information as process to information as content) is more than questionable, but suppose it is accepted; misinforming becomes now a way of informing and misinformation a type of information. All this is as odd as considering lying a way of telling the truth about something else and a contingent falsehood a type of truth on a different topic. The interpretation becomes perfectly justified, however, if informing/information is used to mean, more generically, communicating/communication, since the latter does not entail any particular truth value. But then compare the difference between: (a) "Q is told that p " and (b) "Q is informed that p ", where in both cases p is a contradiction. (a) does not

have to entail that p is true and hence it is perfectly acceptable, but (b) is more ambiguous. It can be read as meaning just “Q was made to believe that p ”, for example, in which case information is treated as synonymous with (a form of) communication (this includes teaching, indoctrination, brain-washing etc.) as presumed by FI.9. But more likely, one would rephrase it and say that (b) means “Q is misinformed that p ” precisely because p is necessarily false, thus implying that it makes little sense to interpret (b) as meaning “S has the information that p ” because a contradiction can hardly qualify as information (...) and being informed, strictly speaking entails truth.” (Floridi, 2005a, pp. 363-364 (and 2011b, p. 96)).

Three points can be made in connection to this quote. First, Floridi acknowledges that Fox’s interpretation is justified if it is about communicating and communication. Fox does not define informing as communicating and information as communication per se, however, his theory is cast in terms of language and communication. Information is what is carried by utterances, contained in sentences, and conveyed through communication. Thus, information is bound by language and communication (cf. section 3.4). Such a notion of information bound by language defined in terms of sentences that express propositions is different from the objective and informee-independent notion that Floridi argues in favor of. Thus, conclusions about the former cannot be imported to serve as opposition to the latter – as that which the latter argues against – without a consideration of the specific contexts wherein the two different theories are developed. As Carnap (1950) points out, the exactness and fruitfulness criteria for the evaluation of an explicatum must be applied within a specified context and established framework (cf. section 1.4). Actually, Floridi (2011b) makes the same point about the need for a specification of the context when he stresses the importance of the CLP parameters (i.e. context, LoA, and purpose) for the Correctness Theory of Truth (CTT) (cf. section 4.2.5) and states that “[i]t is a fallacy to fuse two or more instances of semantic information into a large instance without making their CLP parameters homogenous, at least implicitly.” (Floridi, 2011b, p. 204).

As Fox applies all his conclusions about information to the notion of misinformation with the only difference that misinformation has a fixed truth-value (i.e. it is false) it follows that misinformation is also bound by language and communication. Misinformation is also propositions expressed through sentences which are heard and understood – it is just all the ‘known-to-be-false’ sentences and propositions. Thus, on Fox’s account both information and misinformation are bound by language and communication and they are not direct opposites as information is not required to be true. Instead misinformation is a kind (a subset) of information. As previously argued (cf. section 3.4), Floridi’s notion of information is not bound by language and communication. Furthermore, information and misinformation are defined as direct opposites – i.e. the well-formed, meaningful, and truthful data as opposed to

the well-formed, meaningful, and false data. However, when he defines misinformation, Floridi (2005b) explicitly refers to Fox: “When *semantic content* is *false*, this is a case of *misinformation* (Fox [1983]).” (Floridi, 2005b, section 3.2.3). Thus, in principle Floridi’s strictly semantic notion of information independent of informees is dichotomously opposed to misinformation as a language and communication dependent notion determined by both semantic and pragmatic meaning and dependent upon a misinformee who hears and understands the sentences through which the false propositions are expressed. However, a notion of misinformation dependent upon misinformees who ‘hear and understand’ is not the notion of misinformation entailed by Floridi’s definition of misinformation as well-formed, meaningful, data that is false. This notion of misinformation mirrors the semantic notion of information except for the difference in alethic value.

Second, as previously argued (cf. section 4.2) Floridi’s objection presupposes a notion of information as inherently truthful as well as a notion of ‘inform’ as factive. Misinformation as a kind of information is only ‘odd’ if it is required that information be truthful. Fox would not disagree that “(b) “Q is informed that *p*”, where (...) *p* is a contradiction” (Floridi, 2005a, pp. 363-364) – i.e. cases where *p* is false – could be rephrased in terms of misinforming. However, on Fox’s account the possibility of rephrasing in terms of ‘misinforming’ does not pose a problem as misinforming is a kind of informing – it is the kind of informing where *p* (presumably in retrospect) is known to be false. Interestingly, although neither Floridi nor Fox defines misinformation in terms of misleadingness, if misleadingness as a feature of and requirement for misinformation is taken into account (cf. Fallis, 2014, 2015) a contradiction would hardly qualify as misinformation. A contradiction is no more misleading than a tautology is revealing. The falseness of contradictions¹²⁷ seems as obvious as the falseness of ‘bald-faced lies’ which are not deceiving and misleading precisely due to their obvious falsity (cf. section 1.1.1).

Third, Floridi states that the ninth bad reason for holding that ‘false information’ is genuine information FI.9, presumes an understanding of “Q was informed that *p*” in terms of “Q was made to believe that *p*”. However, Fox (1983) explicitly states that “it is *not* the case that sentences of the form ‘X informs Y that P’ entail the corresponding belief sentences ‘X believes that P’” (Fox, 1983, p. 166) and “it is *not* the case that sentences of the form ‘X informs Y that P’ entail the corresponding belief sentences ‘Y believes that P.’” (Fox, 1983, p. 168). Thus, on

¹²⁷ It can be questioned whether contradictions are always necessarily false, for example, within physics. A famous contradiction which shows some of the paradoxes of physics and quantum mechanics is Schrödinger’s Cat – the cat which is simultaneously dead and alive.

Fox's account belief is neither required on the part of the informer, nor on the part of the informee. As Fox argues that "'inform' and 'misinform' behave alike in all respects except for their relationship to truth" (Fox, 1983, p. 169) it follows that 'misinform' does not require 'belief' on the part of misinformer and misinformee either. Thus, on Fox's account someone can be misinformed without being misled – the misleading depends on whether the misinformation is believed by the misinformee.

I agree with Fox (1983) that information is alethically neutral. Information (that is nonnatural information) has to capture the pragmatic features of communication. These features, for example speaker-meaning, are part of the content and thus part of the information. Sometimes, what is literally said is inaccurate or downright false (e.g. sarcasm), but what is implicated is true and as a whole it is non-misleading. Such utterances (e.g. sarcastic utterances) are information, however, if information is required to be true, the notion cannot capture such utterances as they are inaccurate or literally false. I do not agree with Floridi and Fox that misinformation is necessarily false. I define misinformation as unintended misleading by inaccurate representational content (i.e. by inaccurate information). This definition allows for true misinformation which is the result of true disinformation when the intention to mislead is not preserved through dissemination. True misinformation is also possible as the result of a (putative) implicature – that is, when an agent infers an 'implicatum' where none was intended. Thus, I argue in favor of alethic neutrality for information, misinformation, and disinformation.

6.3 Verbal dispute? Information and truth

Truth in itself is not enough to guard against misinformation and disinformation in order to avoid undesired consequences. As seen throughout the previous chapters Fallis (2009, 2011, 2014, 2015) develops a conceptualization that allows for true disinformation. The notion is derived from the philosophical literature on lying, misleading, and deceiving (Adler, 1997; Fallis, 2010; Mahon, 2008; Stokke, 2013; and Webber, 2013; cf. section 1.1) and therefore cannot easily be dismissed. Fallis' analysis can be extended such that misinformation also can encompass a true variety (cf. section 5.2). Hence, information, misinformation, and disinformation can all be truthful and, if a notion of information as alethically neutral is endorsed, they can all be false as well.

Therefore, truth and falsity are not sufficient conditions in order to differentiate between information on the one hand, and misinformation and disinformation on the other hand. That is, the mere detection of truth and falsity is not sufficient in order to detect information, misinformation, and disinformation. It does not capture 'non-misleadingness' and 'misleadingness', which seems to be the features that most clearly distinguish information from misinformation and disinformation (cf. section 5.3). However, the truth-requirement for

information is still one of the main issues when notions of information are discussed within philosophy of information and the dispute about whether information is veridical or alethically neutral is ever present (cf. Ch. 4). Thus, a conceptual analysis of this dispute, that is, an analysis of whether it is a substantial or a verbal dispute, might provide further insights as to why ‘truth’ and ‘falsity’ in themselves are not sufficient for the detection of information, misinformation, and disinformation, respectively.

The analysis of whether the dispute about information and truth is verbal or substantial serves two goals. As Chalmers (2011) states

“[t]he philosophical interest of verbal disputes is twofold. First they play a key role in philosophical method. Many philosophical disagreements are at least partly verbal, and almost every philosophical dispute has been diagnosed as verbal at some point. Here we can see the diagnosis of verbal disputes as a tool for philosophical progress. If we can move beyond verbal disagreement to either substantive agreement or to clarified substantive disagreement, then we have made progress. (...) Second, verbal disputes are interesting as a subject matter for first-order philosophy. Reflection on the existence and nature of verbal disputes can reveal something about the nature of concepts, language, and meaning.” (Chalmers, 2011, p. 517).

As concepts, language, and meaning are that which algorithms have to analyze in order to detect misinformation and disinformation on the internet, the examination of the dispute can provide some further insights about the notions of information, misinformation, and disinformation and which features most clearly distinguish these three notions when they are detected for.

Through examinations of various proposals for definitions of a verbal dispute Chalmers (2011) arrives at two definitions that define two different kinds of verbal disputes. The first definition is of a narrow verbal dispute and states that:

“A dispute over S is verbal iff S expresses distinct propositions p and q for the two parties, so that one party asserts p and the other denies q, and the parties agree on the truth of p and q.” (Chalmers, 2011, p. 519).

The second is a definition of a “a broadly verbal dispute [which] is one in which an apparent first-order dispute arises *in virtue* of a metalinguistic disagreement. That is:

A dispute over S is (broadly) verbal when, for some expression T in S, the parties disagree about the meaning of T, and the dispute over S arises wholly in virtue of this disagreement regarding T.” (Chalmers, 2011, p. 522).¹²⁸

Although Chalmers (2011) agrees that there might be cases of narrow verbal disputes, he argues that most often a verbal dispute is broadly verbal as it is most often the meaning of some expression that the parties disagree about and not distinct propositions:

“For example, suppose that A and B agree that Sue made a false statement that she did not believe to be false and also agree on the moral status of Sue’s assertion and other relevant properties. A says, ‘Sue did not lie’. B initially says, ‘Sue lied’, believing falsely that ‘lie’ refers to any false statement, but on reflection comes to accept ‘Sue did not lie’, through reflection on the concept of lying. Then, initially, A and B need not be having a narrowly verbal dispute: both may use ‘lie’ to express the same concept. But they are having a broadly verbal dispute all the same: intuitively, they agree on the important facts of the case and are merely disagreeing on whether the word ‘lie’ should be used to describe it.” (Chalmers, 2011, pp. 520-521).

Chalmers’ framework for the treatment of verbal disputes have a Carnapian flavor in the sense that it all comes down to the specification of terms and notions in specific contexts with the overall objective of the fruitfulness or the use of the different specifications of the terms (explicata in Carnap’s vocabulary). This resemblance is acknowledged by Chalmers in his ‘Carnapian Conclusion’ where he writes:

“We also have a Carnapian pragmatism about conceptual frameworks. On this view, instead of focusing on existing words and the concepts they express, we should focus on the role one needs them to play. (...) on his [Carnap’s] project and on this one, the philosophical interest of most concepts lies in the work that they can do for us.” (Chalmers, 2011, p. 564).

¹²⁸ Chalmers (2011) notes that it is “an *explanatory* “in virtue of”: the idea is that the metalinguistic disagreement *explains* the apparent first-order disagreement.” (Chalmers, 2011, p. 525).

6.3.1 Methods for resolving verbal disputes

According to Chalmers (2011) “a broadly verbal dispute is one that can be resolved by attending to language and resolving metalinguistic differences over meaning.” (Chalmers, 2011, p. 526). One approach is to

- settle the facts about key terms in the context, and then
- to distinguish senses of the key terms.

Another approach is the *method of elimination* which falls in three steps:

“First: one bars the use of term T. Second: one tries to find a sentence S’ in the newly restricted vocabulary such that the parties disagree nonverbally over S’ and such that the disagreement over S’ is part of the dispute over S. Third: if there is such an S’, the dispute over S is not wholly verbal, or at least there is a substantive dispute in the vicinity. If there is no such S’, then the dispute over S is wholly verbal (...).” (Chalmers, 2011, p. 527).

Thus, in order to settle whether or not a dispute is verbal the parties who disagree over some term T can be asked to restate the dispute or parts of it without the use of term T (or someone can try to restate the dispute for them). If the dispute can be restated and the new dispute is nonverbal, then the initial dispute was substantial. However, if the dispute cannot be restated without the use of term T, or if the dispute can be restated but the new dispute is verbal, then the initial dispute was verbal as well. Furthermore, it can be asked what turns on the dispute: “If something turns on the verdict it can still be a verbal dispute, but it is not a merely verbal dispute.” (Chalmers, 2011, p. 525).

6.3.2 The dispute – Information and truth

The present dispute is the one in regard to information and truth. Floridi (2005a, 2005b, 2007, 2011b) argues in favor of a claim, which can be phrased as

(1) Information requires truth in order to be information

whereas Fallis (2009, 2011, 2014, 2015), Fetzer (2004a), and Fox (1983) argues in favor of a claim, which can be phrased as

(2) Information does not require truth in order to be information.

In order to detect whether this dispute is verbal or substantive the term information must be barred from the sentences (1) and (2), as described in the method of elimination. When ‘information’ is barred, (1) and (2) can be restated in the new restricted vocabulary. As the dispute concerns the definition of the concept information Floridi, Fallis, Fetzer, and Fox

provide descriptions of ‘information’ which can be used to restate (1) and (2). In order to restate (1) Floridi’s definition of information is used and in order to restate (2) Fallis’ definition of information is used (as it also captures Fox’s and Fetzer’s definitions). Thus, (1’) and (2’) are generated:

- (1’) Well-formed, meaningful, truthful data requires truth in order to be well-formed, meaningful, truthful data.
- (2’) Representational content does not require truth in order to be representational content.

However, these two restatements are both necessarily true. It cannot be denied that something truthful requires truth in order to be truthful and it cannot be denied that representational content as such does not require truth in order to be representational content. Thus, the two parties have to agree on the truth of (1’) and (2’) but that does not resolve the initial dispute. However, it is clear that the dispute concerns the meaning of the term information – i.e. what is implied by that concept and thereby when it can be applied. Although it is a dispute which concerns the meaning of the term information, the dispute is different from the one in the example about ‘lying’. It is not the case that one of the parties (i.e. Floridi vs. Fallis, Fetzer, and Fox) will suddenly realize that they, in accordance to the definition of the term, applied the term wrongly. The dispute about information and truth is a dispute about the definition of the term or concept information. Thus, the parties cannot come to an agreement about a ‘correct’ application of the term information, as there is no agreement on the term information.

Therefore, it might be more fruitful to state the dispute as a classic philosophical question of the kind ‘What is X?’ – i.e ‘What is information?’ – because that is actually what the dispute is about. According to Chalmers (2011) disputes of the kind ‘What is X’ are almost always verbal. However, that does not mean that it cannot provide philosophical progress to resolve these disputes as a clarification of verbal disagreement is progress in itself.

Thus, the dispute between Floridi and Fallis, Fetzer, and Fox concerns the question ‘What is information?’ Floridi’s answer to that question is

- (3) ‘Information is well-formed, meaningful, and truthful data’, (i.e. truthful semantic content)

whereas Fallis’ answer (as a proponent for Fetzer and Fox as well) is

- (4) ‘Information is representational content’ (i.e. alethically neutral).

When the disputed question is of the kind 'What is X?' there is a special case of the method of elimination which can be applied, the so-called subscript gambit (Chalmers, 2011). In this case of the method the disputed term T (i.e. information) is barred in order to introduce two new terms Information₁ and Information₂ stipulated to be equivalent to the two right-hand sides. Hence, (3) and (4) are restated as

(3') 'Information₁ is well-formed, meaningful, and truthful data, (i.e. truthful semantic content).'

and

(4') 'Information₂ is representational content (i.e. alethically neutral).'

If the parties have nonverbal disagreements which involve Information₁ and Information₂ then the initial dispute is most likely nonverbal as well (i.e. it is substantive) or at least has substantive elements. However, if the parties do not have nonverbal disagreements which involve Information₁ and Information₂ then the initial dispute is verbal. In the case of (3') and (4') there is no nonverbal dispute left, wherefore the dispute is verbal. Fallis would not deny that there is a subset of well-formed, meaningful data (i.e. semantic content) which is truthful – that is Information₁ – he just denies that this subset is the only thing which counts as 'information'. Likewise, Floridi acknowledges that a concept of semantic content in general is needed to capture both information, misinformation, and disinformation and this is exactly what Fallis' notion of representational content – i.e. Information₂ – does. Floridi just denies that this general concept of semantic or representational content should be called information. Thus, Floridi and Fallis agree that Information₁ is a subset of Information₂ but they do not agree which of the two concepts should be termed 'information'. For Floridi (2005a, 2005b) semantic content in general (i.e. well-formed, meaningful data) is divided by truth and falsity into information (the truthful part) and misinformation and disinformation (the false part). Fallis (2009, 2011, 2014, 2015), on the other hand, sees misinformation and disinformation as kinds of information and he does not divide representational content by truth and falsity as he wants to capture the notion of true disinformation.

6.3.3 What turns on the dispute?

Although the dispute as to whether information requires truth or not is verbal the possibility of true disinformation and true misinformation does turn on the outcome of the dispute. As seen in Ch. 1, Ch. 3, and Ch. 5, Fallis' notion of true disinformation and the development of the notion of true misinformation (cf. section 5.2) are based on the notion of false implicature in line with the literature on lying, misleading, and deceiving (cf. section 1.1).

Floridi (2005a, 2005b) acknowledges that semantic content in general (i.e. well-formed, meaningful data) is needed in order to capture information, misinformation, and disinformation, alike, as different kinds (semantic content in general is that which information, misinformation, and disinformation are kinds of) – that is, he accepts Information₂ as a necessary concept which, when truth-values are added is either information (truthful) or misinformation and disinformation (false). However, as Floridi accepts semantic content as capturing information, misinformation, and disinformation, alike, but distinguishes these notions by truth and falsity, he ends up in a position where he must either

1. deny the possibility of true misinformation and true disinformation, or
2. accept that true misinformation and true disinformation are captured by his definition of information as ‘well-formed, meaningful, and truthful data’.

To deny the possibility of true misinformation and true disinformation would demand that the body of literature on lying, misleading, and deceiving be rewritten as the notion of false implicature, as the means for verbal misleading and deceiving, is more or less uniformly agreed upon. Thus, the truth-false dichotomy for information vs. misinformation and disinformation collapses. As it is unspecified on Floridi’s account, that information must be non-misleading, true misinformation and true disinformation are captured by Floridi’s notion of semantic information. It seems that within philosophy of information the defenders of information as necessarily truthful assume that because information is true per definition it cannot be misleading – information is always good. It is the true ingredient in knowledge and it secures knowledge better than justification. However, when true misinformation and true disinformation are possible this view is problematic or at least disputable. The literal truth of true misinformation and true disinformation in itself is not a guarantee for non-misleadingness – i.e. a guard against misleadingness – as the disinformation and the misinformation still are capable of generating false beliefs on the part of the disinformee and misinformee. If ‘truth’ was enough to guard against misleadingness true misinformation and true disinformation would not be possible.

As mentioned, the dispute between Floridi and Fallis in regard to the truth-requirement for information is a verbal dispute which arises in virtue of a metalinguistic disagreement over the term ‘information’. There is no substantial disagreement in the sense, that Floridi and Fallis agree on the ‘existence’ of two different terms Information₁ and Information₂. The disagreement concerns whether it is Information₁ or Information₂ which earns the label ‘information’. The outcome of the dispute is another matter. To settle whether a dispute is verbal or substantial does not necessarily settle the dispute itself. However, the clarification of the metalinguistic disagreement over the term ‘information’ has the potential to “move [the disagreeing parties] beyond verbal disagreement to either substantive agreement or to

clarified substantive disagreement” (Chalmers, 2011, p. 517) which, according to Chalmers, is philosophical progress in itself. Thus,

“[T]he method I have outlined has the potential to clarify many or most philosophical disputes, and to resolve some of them. It will certainly not resolve them all, but it often gets us closer to the heart of the dispute and leaves us with a clearer understanding of the fundamental issues underlying a dispute and of what remains to be resolved. That is a form of philosophical progress.” (Chalmers, 2011, p. 564).

What remain to be resolved are some further aspects of information, misinformation, and disinformation and their interconnections before a unified conceptualization is possible.

6.4 Disinformation as a success term?

Floridi (2005b, 2011b) acknowledges that disinformation is purposefully or intentionally conveyed as opposed to misinformation which is unintentional (Floridi, 2005b). However, as there is no specification of the intentionality (in the directedness-/foreseeability-sense) or unintentionality of information ‘well-formed, meaningful, and truthful data’ still captures true misinformation and true disinformation. One feature which could exclude true disinformation from ‘well-formed, meaningful, and truthful data’ – i.e. semantic information – is if disinformation is a “success term”. If disinformation is a success term it means that something is only disinformation if someone is actually misled – i.e. if the misleading is successful. As Floridi’s (2005a) notion of semantic information is independent of informees (in the sense of someone receiving the information) it cannot capture a notion of disinformation as a success term as such a term presupposes a receiver and thus is dependent on a receiver who is misled.

According to Fallis (2015) disinformation is not a success term, that is, even though no one is actually misled by the intentionally misleading representational content it is still disinformation and in the act of providing it one has disinformed. Thus, disinformation is independent of informees (i.e. disinformees) and the true variety is still captured as an instance of well-formed, meaningful, and truthful data along with true misinformation. On Fallis’ (2014) account of disinformation lies are a type of disinformation¹²⁹ and disinformation is a kind of deception. According to Mahon (2008) ‘to lie’ is not a success term but ‘to deceive’ is an achievement or success term (cf. section 1.1). This means that lies and

¹²⁹ Except ‘bald-faced lies’ which are not intended to mislead and are not intentionally misleading either, as they are not misleading at all (cf. Fallis, 2014, note 4).

disinformation are not always deceptive – i.e. they are not always instances of deception. Lies and disinformation can be deceptive and when someone is actually misled by the lie and the disinformation then it is deception. If no one is misled then the lie and the disinformation are not instances of deception but they are still a lie and disinformation.

6.5 A unified conceptualization¹³⁰

‘A unified conceptualization’ refers to a conceptualization of information, misinformation, and disinformation as a whole. The conceptualization is developed on the basis of the interconnections of information, misinformation, and disinformation and all three notions are attended to equally. It is not first and foremost an account of information, or of misinformation, or of disinformation. It is an account of all three notions together. Thus, ‘a unified conceptualization’ does not refer to a conceptualization where all the different explicata examined and discussed throughout the previous chapters are encompassed. From the analyses and discussions of the various explicata through the aspects of ‘meaning’, ‘truth’, and ‘intention’ it is seen that the different explicata are explicata of different explicanda. They are developed within different contexts and with different purposes – i.e. they are parts of different projects – wherefore they does not necessarily fit into a single unified conceptualization.

6.5.1 Explicanda and explicata

Dretske’s (1981) emphasis is on cognition and learning with the outset of so-called Information-Theory (i.e. the mathematical theory of communication) which is one of the more technical explicanda of information (cf. section 1.2.2). Dretske’s (1981, 1988) project is to naturalize intentionality in order to explain how meaning, minds, and cognition arise, to offer an informational account of knowledge, and to explain behavior. Within Dretske’s development of semantic information a notion of misrepresentation, which can lead to misinformation, occurs. Dretske’s explicatum of misinformation is ‘false semantic content’ which bears some resemblance to the explicanda of misinformation as ‘false information’ if ‘information’ in this context is understood simply as semantic content and not as Dretske’s own notion of information.

Floridi’s (2004b, 2005a, 2005b, 2007, 2011b) notion of information is developed as the center of his ontological proposal for a philosophy of information which encompasses a constructionist view of knowledge. Floridi’s (2002b, 2004a, 2011a, 2011b) project concerns

¹³⁰ Minor parts of this discussion have been put forth in Sørensen (submitted) *Intention and Misleading*.

the semanticization of the universe, more specifically, how meaning is acquired and ascribed – it is the idea that the universe and everything in it can be explained in informational terms. The aspect of semanticization on Floridi's account resembles Dretske's purpose of naturalization of intentionality, however, their approaches and thus their explicata are quite different. Floridi's explicatum of information does not bear any striking resemblance to any of the explicanda. Floridi's notion of semantic information is connected to knowledge, it is that which secures knowledge, but it is not linked to language and communication. Instead, it is partly developed from an Information-Theoretic notion of information but there is not much similarity between Shannon's quantitative notion of information and Floridi's semantic notion of information. On the other hand, the similarity between the explicanda and the explicata of misinformation and disinformation on Floridi's account is striking insofar as the references to 'information' in the explicanda are taken to mean semantic content in general. Although the similarity between explicanda and explicatum on Floridi's account of information is small, the relevant explicanda (i.e. an information-theoretic notion of information and a connection to knowledge) are the same as the explicanda which are Dretske's outset for a theory of semantic information.

Fox's (1983) explicatum of information is developed within the context of communication and is very similar to the explicandum of information as 'a message received and understood'. This explicandum is very far from the explicanda of information as connected to knowledge or as information-theoretic quantities. Fox's purpose is to capture and provide an analysis of the 'ordinary notion of information' and he explicitly states why information-theory is not suitable for the pursuit of this task:

"Even a proponent of the position that information theory is relevant to the philosophical analysis of information (namely Fred Dretske), admits that

communication theory does not tell us what information is. It ignores questions having to do with the *content* of signals, what *specific information* they carry, in order to describe *how much* information they carry (Dretske, 1981, 41).

Dretske could have added that communication theory also ignores questions having to do with the beliefs, intentions, reliability, and other features of the informant and informee in processes of informing.

In neglecting so many crucial issues, information theory forfeits any chance of being a legitimate candidate for providing an adequate analysis of the ordinary notion of information. (Of course, this was never the goal of information theory anyway.)" (Fox, 1983, pp. 55-56).

Although Dretske also states that he wants to capture the ordinary notion of information it is two very different 'ordinary notions' which Dretske and Fox have in mind and they are derived from, or are similar to, two different explicanda. Fox's explicatum of misinformation is similar to the explicandum of misinformation as 'false information' and the relation or connection between information and misinformation resembles the relation or connection from Vocabulary.com where misinformation is a type of information in the 'message-received-and-understood' sense.

Fallis' (2009, 2011, 2014, 2015) explicatum of information as 'representational content' in general is developed within the scope of a conceptual analysis of disinformation – that is, the explicatum of information primarily serves to help explain and qualify the explicatum of disinformation. Disinformation is explicated as representational content which has the function to mislead, where 'function' entails intentionality in the directedness- and foreseeability-sense. This explicatum of disinformation resembles the explicanda in terms of 'intentional misleadingness' but is more exact as it considers all the different varieties of disinformation including visual disinformation, side effect disinformation, and true disinformation. Fallis' explication of information is more similar to an interpretation of information as found in the explicanda of misinformation and disinformation, than to the actual explicanda of information. The explicanda of misinformation and disinformation refer to a notion of information, which can be misleading and inaccurate, in order to define misinformation and disinformation. It seems to be this notion of information which Fallis explicates into 'representational content', rather than a notion of information related to knowledge, learning, and the like. That Fallis' explicatum of information is similar to a notion of information within the explicanda of misinformation and disinformation and not to any explicandum of information might stem from the fact that information is explicated as part of the explication of disinformation. 'Information' enters the conceptual analysis of disinformation as an explanatory feature which is not up for analysis itself. Fallis' explicatum of misinformation is similar to the explicandum of misinformation as 'inaccurate and misleading information'.

Scarantino and Piccinini's (2010) explicata of information as 'natural information' and 'nonnatural information' do not seem to have the explicanda from the dictionaries as their explicanda. Instead, the double-account of information is developed as a further explication of the explicata of information provided by Dretske and Floridi, respectively, and with Grice's notions of natural and nonnatural meaning as additional explicanda. Thus, the explicatum 'natural information' is a further explication of Dretske's notion of semantic information, whereas the explicatum 'nonnatural information' is developed as an alternative to Floridi's notion of semantic information. Scarantino and Piccinini cast their explications within the context of cognitive science and computer science and argue that their explicata – especially

nonnatural information – are more fruitful as they capture the way cognitive scientist and computer scientist speak of, and work with, information, for example in order to explain information-processing in the brain. Scarantino and Piccinini dismiss a notion of misinformation as different in kind from information and generally prefer to speak of ‘false (nonnatural) information’ – an explicatum which is similar to an explicandum of misinformation as ‘false or misleading information’.

The evaluation and comparison of the various explicata of information, misinformation, and disinformation against the various explicanda of the three notions show that although two theories explore the same explicandum of information they might end up with completely different explicata (e.g. Dretske and Floridi). It also shows that although two theories set out with the purpose of capturing the ‘ordinary sense’ of information, they might choose two different explicanda as their outset and therefore develop different explicata (e.g. Dretske and Fox). Furthermore, as Carnap (1950) points out with his requirement for fruitfulness within a specified context, the overall purpose of the explication determines the choice of explicandum which then determines which explicata are possible. For instance, the naturalization of intentionality and the explanation of how meaning and minds arise cannot depart from a notion of information defined in terms of knowledge, learning, and the like, as such an explicandum presupposes the concepts which the explicatum is supposed explain the origin of. Thus, in the context of automatic detection of information, misinformation, and disinformation in online social network structures the explicata developed within a context of language and communication offer themselves as more fruitful.

However, none of the accounts of information, misinformation, and disinformation examines the interconnections between these three notions. When the interconnections are examined, analyzed, and discussed it shows that none of the explicata are fruitful for automatic detection as they are not developed and defined in accordance to one another – i.e. in accordance to the interconnections. Thus, of the present explicata it is not possible to pick three (one of information, one of misinformation, and one of disinformation) and combine them in a unified conceptualization. However, such a unified conceptualization is exactly what is needed in connection to automatic detection as it has to be clear what is detected for and why.

6.5.2 The conceptualization

I have argued that the distinction between information vs. misinformation and disinformation in terms of truth and falsity collapses with the entrance of true disinformation – especially when this kind of disinformation is captured by Floridi’s notion of semantic information. Instead I propose ‘non-misleadingness’ as a determinant feature of information, which is then distinguished from misinformation and disinformation as ‘misleading’. In order to create a unified conceptualization of information, misinformation, and disinformation some

discussions of how information as non-misleading is connected to misinformation (as unintended misleading) and disinformation (as intentional misleading) still remain.

There are two basic approaches that can be applied: Either 'information' is an overall term which encapsulates both 'misinformation' and 'disinformation' as varieties of information, or 'information', 'misinformation', and 'disinformation' are distinct notions connected in a linear way, where none of the notions are varieties of each other.

If the first approach is taken and information is an overall term, then information cannot be defined as 'non-misleading'. If information is 'intentionally non-misleading', then disinformation as 'intentional misleading' cannot be a variety of information. Something which is 'intentionally non-misleading' cannot encapsulate something which is 'intentionally misleading' – that would be a contradiction. If misinformation and disinformation are varieties or subcategories of information, then information is defined as everything with representational content – misleading and non-misleading, true and false i.e. alethically neutral – just as Fallis (2014) proposes. However, when misinformation is defined as 'unintended misleading', regardless of its truth-value (which means that misinformation is all that which has representational content and is unintended misleading) and disinformation is defined as 'intentional misleading' regardless of truth-value (which means that disinformation is all that which has representational content and is intentionally misleading), then all that which is 'intentionally non-misleading' seem to disappear or is at least not explicit anywhere. As mentioned, 'non-misleading' cannot be a definitional part of information because then information can no longer encapsulate misinformation and disinformation as subcategories¹³¹.

If the second approach is taken and information, misinformation, and disinformation are all distinct notions in the sense that they are not varieties or subcategories of each other, then information can be defined as 'intentional non-misleading'. In this approach the three notions are unified as three distinct notions on the same level – i.e. they are not varieties of one another – information is 'intentional non-misleading', misinformation is 'unintended misleading' but was intended to be non-misleading, and disinformation is 'intentional misleading'. In this way representational content (e.g. a sentence, utterance, picture, gesture, etc.) can be mapped according to whether or not it is misleading and whether or not the misleading/non-misleading is intentional/intended. If the intentionality or intention as well as the misleadingness/non-misleadingness of *p* are not directly preserved through

¹³¹ Disinformation cannot be a subcategory of misinformation either as something intentionally misleading cannot be a variety of something unintended misleading.

dissemination but only indirectly preserved through transitivity shifts between the three notions disinformation, misinformation, and information are possible (cf. section 5.3). The intentionality/intention is dependent on the agent (the disseminator), whereas the misleadingness/non-misleadingness is dependent on the context (e.g. the frame, the situation, etc.) in connection to the accuracy or inaccuracy of p as well as the receiver.

Accuracy and inaccuracy come in degrees and are in themselves neither true nor false. Inaccuracy is not in itself misleading, although it can be as in the example of clinical trials (cf. section 5.3). Grice's notion of implicature (1967) is a fine example of the alethic neutrality of inaccuracy. When implicatures are used what is said is inaccurate albeit not necessarily false and, in general, not misleading. However, implicatures can be misleading if what is implicated is false (i.e. false implicatures), as in the case of true disinformation (Fallis, 2014; cf. sections 1.5.2, 2.3.3, and 3.5.1) and true misinformation (cf. 5.2). The misleadingness or non-misleadingness of implicatures depends on the truth-value of what is implicated rather than the inaccuracy of what is said.

The definition of information as intentional non-misleading does not entail truthfulness as a requirement, which means that information can be alethically neutral. Furthermore, it is worth noting that Fox (1983) and Fallis (2009, 2011, 2014, 2015) –who work explicitly with the notions of misinformation and disinformation, respectively – both adhere to a notion of information as alethically neutral. In contrast Dretske (1981, 1983, 2008) and Floridi (2004b, 2005a, 2005b, 2007, 2011b) – who almost exclusively work with the concept of information – both adhere to a notion of information as necessarily veridical (cf. section 3.2 and Ch. 4). This seems to indicate that the notion in which an analysis takes its departure determines how this notion, as well as the other notions, can be conceived – i.e., as Floridi would say, the context, the LoA, and the purpose determine the question and thus determine the answer.

The adherence to information as alethically neutral in order to define misinformation and disinformation as well as categorize them as varieties of information, which is what Fox (1983) and Fallis (2009, 2011, 2014, 2015) do, suggests that information as an overall term is not easily abandoned. Add to this the way in which the notion of information is commonly used as a collective notion for what is found on the internet, for example, independently of truth-value (i.e. representational content in general), a use which is also indicated by some of the dictionaries (cf. section 1.2) and is adhered to by the detecting-projects.

Thus, a combination of the two approaches to the unification of information, misinformation, and disinformation and thereby two different notions of information is desired. These two notions of information are

1. Information as an overall term i.e. representational content in general, which includes misinformation and disinformation as varieties, and also includes
2. Information as intentional non-misleading.

However, this conceptualization does not capture the distinction between natural and nonnatural meaning in terms of natural and nonnatural information. Although the scope is to evaluate the various explicata in accordance to their fruitfulness for automatic detection in online social network structures based on short semantic structures (i.e. short written texts) – that is, structures bound by language as part of communication – natural meaning and natural information, for example as a way to explain cognition, cannot simply be left out. Natural information is needed in order to naturalize intentionality (Dretske, 1981; Jacob, 2003; and Scarantino & Piccinini, 2010) – nonnatural information cannot do that as it presupposes intentional content, i.e. intentionality in the aboutness-sense:

“[A] critic may interject that the notion of nonnatural information is problematic, because it presupposes intentional content (say, the content *that the bus is full*). Because of this nonnatural information cannot be used to naturalize intentionality, which is one of the central projects that have attracted philosophers to information in the first place (Dretske, 1981, 1988;...). (...) We agree that nonnatural information is in need of naturalistic explication just as much as intentional content. (...) But our present goal is not to naturalize intentionality. Rather, it is to understand the central role played by information in the sciences of minds and in computer science.”
(Scarantino & Piccinini, 2010, p. 315).

As argued by Scarantino and Piccinini (2010) the distinction between natural and nonnatural information marks the distinction between physical occurrences in the world and convention, language, and communication. Thus, in the context of automatic detection and agents who act in different ways the focus is on nonnatural information although it is acknowledged that natural information must be part of the unified conceptualization.

Therefore, a new conceptualization must be developed:

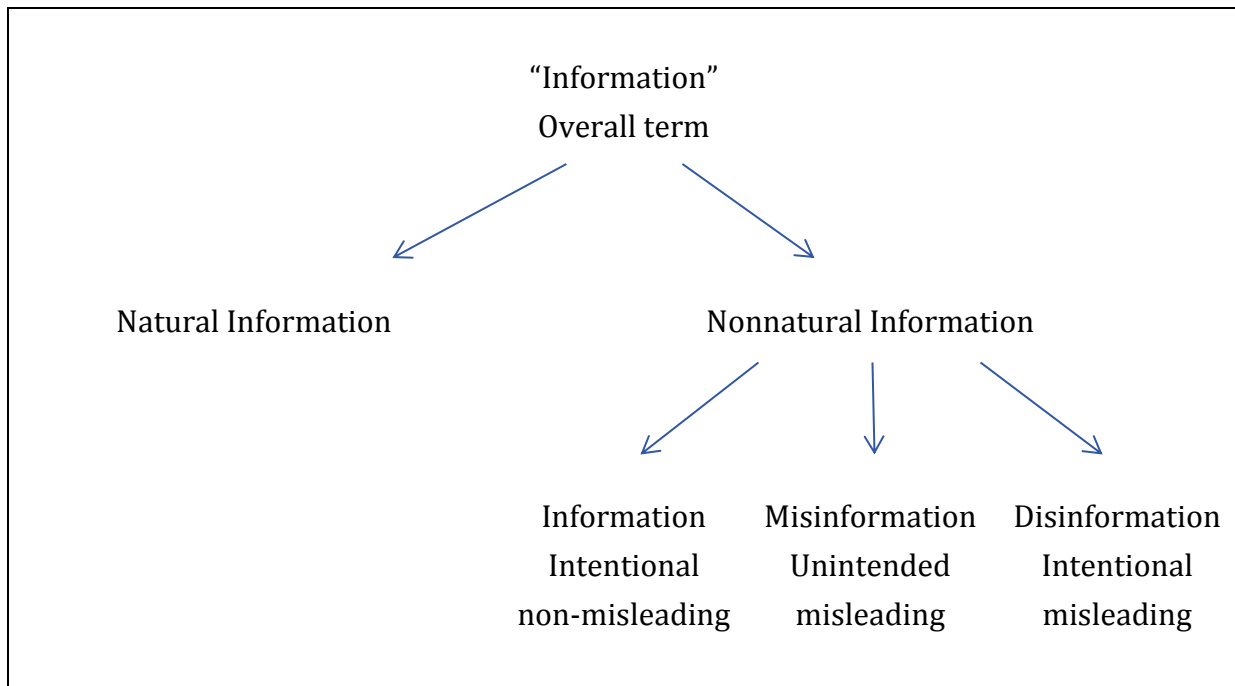


Fig. 6.1 – Three levels and four notions of information.

This unified conceptualization has three levels and four notions of information.

The first level includes one notion of information:

- “Information” – overall term.

This is the overall notion of information which refers to all the different explicanda of information – i.e. the concept which is used, when some specific notion of information is not specified. It is the neutral mass noun, the conglomerate of all the different shapes and interpretations of information which can be meant when people just say ‘information’.

The second level includes two notions of information which are both semantic notions:

- Natural information and
- Nonnatural information which is roughly equivalent to representational content in general.

The third level includes:

- information as intentional non-misleading,
- the notion of misinformation as unintended misleading, and
- the notion of disinformation as intentional misleading.

Thus, information (specified as intentional non-misleading), misinformation, and disinformation are kinds of nonnatural information – they are different aspects of nonnatural information. Disinformation, for instance, is the aspect of nonnatural information where the information is intentionally misleading and can be deceptive. Misinformation is the aspect of nonnatural information where the information is unintended misleading and therefore cannot be deceptive, and information is the intentionally non-misleading aspect.

All four notions of information and the two notions of misinformation and disinformation are alethically neutral. The alethic neutrality of natural and nonnatural information follows from the argument by Scarantino and Piccinini (2010) (cf. section 4.2.3), the alethic neutrality for disinformation follows from Fallis (2009, 2011, 2014, 2015), and the alethic neutrality for misinformation follows from the discussion in section 5.2. Although Dretske (1981, 1986) has a notion of misrepresentation which, when agents enter the picture, enables the generation of misinformation it is more fruitful to reserve the notions of misinformation and disinformation to designate the two misleading aspects of nonnatural information. Misinformation and disinformation have to do with agents, actions, and communication, something which is not present within the definition of natural information. Agents, actions, intentions, beliefs, and the like are closely connected to nonnatural information which is agent-dependent in the sense that it deals with conventions, which are manufactured on the basis of intentions and presuppose intentional content. In connection to natural information there is the possibility of misrepresentation and error. For example, if a thermometer malfunctions and provides the wrong temperature, then it is an instance of misrepresentation and error, but it is not a case of misinformation as the error is not produced unintended by an agent. However, if the agent who reads the thermometer tells someone else the wrong temperature based on the reading of the thermometer, then the utterance is an instance of misinformation and an act of misinforming. Or it is an instance of disinformation and an act of disinforming if the agent knows that the reading on the thermometer is wrong and he disseminates it anyway. When the agent reads the thermometer he forms beliefs about the temperature in the room. When these beliefs are expressed through language they enter the realm of nonnatural information as an instance of communication and thus it can be an instance of misinformation (if it is unintended misleading) or disinformation (if it is intentional misleading). If the reading on the thermometer is correct, or if the agent knows that the thermometer is malfunctioning and thus includes this knowledge in his utterance about the temperature in the room, then the utterance is an instance of information as intentional non-misleading.

Although misrepresentation and error in themselves are not misinformation or disinformation they can still be misleading – the reader of the thermometer can be misled by the incorrect reading, the difference is that the incorrect reading is not provided by an agent, it is provided by the thermometer. Something is misleading if it is likely to cause the receiver

to hold a false belief or likely to prevent the receiver from obtaining a true belief – that is, ‘misleading’ is something which is likely to leave the receiver epistemically worse off than she could have been (Fallis, 2014). Non-misleadingness, on the other hand, is likely to cause the receiver to hold a true belief or likely to get the receiver to dismiss a present false belief – that is, ‘non-misleadingness’ is something which is likely to leave the receiver epistemically better off than before or at least not to leave her epistemically worse off than she could have been.

The connections between information, misinformation, and disinformation across three different levels and with four different notions of information offer an explanation as to why there is no agreement on a single notion of information. When people speak of ‘information’ and when philosophers and information studies scholars discuss ‘information’ they do so on three different levels without clarifying and specifying which level they refer to. Some use of the notion information might be on the first level as an overall term where others respond with a notion from the second level (e.g. natural or nonnatural information) as opposed to notions from the third level (misinformation and disinformation). Thus, according to the unified conceptualization it is correct that misinformation is not a kind of information in Dretske’s (1981, 1983, 2008) sense – i.e. is not a kind of natural information – however, misinformation (also in Dretske’s sense) is a kind of nonnatural information. Although Floridi requires information to be truthful, a requirement which is not endorsed in the unified conceptualization where the notions are defined as alethically neutral, his notion of semantic information can be captured by the notion of nonnatural information¹³². Truth and falsity are not the defining features of information vs. misinformation and disinformation, however, the specific instances of the various notions – i.e. the specific pieces of information, misinformation, and disinformation – still have truth-values. Thus, nonnatural information encompasses both true and false representational content. Floridi’s reference to Fox’s notion of misinformation as directly opposite to his own notion of information is a reference across two levels as Floridi’s semantic information is not captured by the notion of information as intentional non-misleading.

In order to determine whether an instance of nonnatural information (expressed through language or gestures) is intentionally non-misleading information, misinformation, or disinformation Grice’s Cooperative Principle and its maxims offer a heuristic:

- If the Cooperative Principle is observed and no maxims are violated, flouted, exploited, etc. then it is non-misleading information.

¹³² According to Scarantino and Piccinini (2010) Floridi’s notion of semantic information belongs to the concept of nonnatural information.

- If maxims are violated, flouted, exploited, etc. in a way which generates true implicatures then it is also non-misleading information.
- If a maxim is violated, flouted, exploited, etc. unintended and it generates a (putative) false implicature; or it does not generate a true implicature and it is not mere opting out, then it is misinformation.
- If a maxim is violated, flouted, exploited, etc. intentionally and it generates a false implicature; or it does not generate a true implicature and it is not mere opting out, then it is disinformation.

As these four points are heuristics they might not provide the right answer on all occasions and they might be difficult to apply in online network structures where the context is not always apparent. However, as Chapman (2005) points out “the programmes that are most successful are those that adhere most closely to Grice’s maxims” (Chapman, 2005, pp. 195-196)¹³³.

¹³³ For the full quote on Gricean maxims and AI, cf. section 2.5.

Conclusion

What to detect?

This endeavor began with ‘the urge to detect’ misinformation and disinformation, as well as information in online social network structures. Projects such as PHEME (2014), Kumar and Geethakumari’s (2014) algorithm-project, and Karlova and Fisher’s (2013) diffusion model all place emphasis on the potential dangers the spread of misinformation and disinformation in social network structures might cause (cf. Introduction). I have argued that in the context of automatic detection the features which distinguish information, misinformation, and disinformation from one another are non-misleadingness/misleadingness and intention/intentionality – not truth and falsity (cf. Ch. 6). However, to automatically detect these features might prove rather complicated. To determine whether something is true or false can be complicated enough (cf. Ch. 4). The truth and falsity of utterances (both written and oral), for instance, are sometimes dependent on the context within which the utterance is made (who made the utterance, where was it made, when was it made, and in which connection was it made?) – this is the point Grice (1957, 1967) makes when he introduces the distinction between word-/sentence-meaning and speaker-meaning (cf. Ch. 2). The PHEME-project (2014) highlights the importance of the context when the researchers put special emphasis on the detection of the spatio-temporal dimensions of tweets. In the theories of meaning (cf. Ch. 3) the challenge of truth-value ascriptions and the need for pragmatics are recognized in connection to indexicals (e.g. ‘I’, ‘here’, ‘now’, ‘nearby’, etc.) which are unspecified references to specific persons, places, and times, which can only be determined when the context is known.

To determine whether some instance of representational content is misleading or non-misleading seems even more complicated. For instance, ‘misleadingness’ (when not the

product of mere lying) is dependent on false Gricean implicatures, or it is generated by inaccuracy (i.e. the whole picture is not provided), and is partly dependent on the receiver (cf. Ch. 1). Whether something is misleading for some agent is partly dependent on what that agent knows in advance. The provision of the results from four clinical trials which show one effect of a drug on humans may not be misleading for an agent who already knows about the other 50 clinical trials which show another effect of the drug (cf. Ch. 5). For such an agent the four new results might actually provide valuable insights in regard to the 50 clinical trials. However, the results from the four clinical trials are misleading due to inaccuracy if an agent does not know about the 50 clinical trials and their results.

Therefore, it might not be possible to automatically detect misinformation and disinformation, or information. The distinctions are cast in terms of intention/intentionality and misleadingness/non-misleadingness, wherefore these features have to be detected – but how can intentions and intentionality be detected automatically? In connection to ‘intention’ and ‘misleadingness’ various proposals for textual cues to deception have been put forth as suggestions for automatic detection with the note that further research on these cues is needed (Karlova & Lee, 2011). However, according to forensic linguist Robert A. Leonard (2015, personal communication) there is no such thing as textual cues to deception – it is all about having enough context. It is a matter of having enough context in the sense of metadata and plenty of writing samples from the involved parties in order to

- determine which features of some written text are cues to deception in that specific text for that specific writer, or
- in order to determine whether the alleged author of some text is in fact the author and not a fraud.

Thus, it might be more fruitful to detect for something else, for instance trust and credibility. Trust and credibility are in fact what Kumar and Geethakumari’s (2014) proposed algorithm detects. However, the detection of trust and credibility does not eliminate all errors – i.e. it is still possible to be misled (cf. section 5.4.1). Whether one chooses to detect for misleadingness/non-misleadingness and intention/intentionality or trust and credibility the Gricean maxims and the Gricean heuristics (cf. section 6.5.2) might lead the way. For instance, Grice (1967) notes that the infringement of the second maxim of quantity (“Do not make your contribution more informative than is required”) suggests that either the utterer does not actually know what he is talking about, he is not as certain as he appears, or the information he provides is actually quite controversial. However, as long as the misleadingness/non-misleadingness of representational content is partly determined by what the receiver already knows it does not seem possible to automatically detect all instances of non-misleading information, misinformation, and disinformation.

What is clarified?

The investigation of the conceptualizations of information, misinformation, and disinformation, and especially their interconnections, within philosophy of information is initiated by a specification of the field. Philosophy of information is most often associated with philosopher Luciano Floridi and his 'world view' interpretation of the field. Instead of subscribing to Floridi's view, I offer a new interpretation of philosophy of information as an interdisciplinary field comprised of the intersections between the contributing disciplines: computer science, cognitive science, artificial intelligence, (analytical) philosophy, information studies, and the like. This 'intersection-approach' enables contributions from both sides of an intersection as instances of philosophy of information as long as the goal is a philosophical and critical investigation of 'information'. In this thesis it is the intersection between analytical philosophy (epistemology and philosophy of language) and information studies which is emphasized. The intersection-approach is especially fruitful in regard to this particular intersection as the contributions from within information studies are more or less neglected within philosophy (with Floridi as exception). As the intersection-approach to philosophy of information is not present within the literature of the field (except for this thesis), the interpretation is in itself a contribution to philosophy of information.

The explorations, analyses, and discussions of the notions of information, misinformation, and disinformation are conducted in the field of philosophy of information (cf. Ch. 1), within a Gricean framework of meaning and cooperation (Grice 1957, 1967; cf. Ch. 2), and through Carnap's (1950) method of explication (cf. section 1.4).

The conceptualizations of information, misinformation, and disinformation and their interconnections as they are presented by selected philosophers and information studies scholars within philosophy of information are shaped by 'meaning', 'truth', and 'intention' in different ways. Furthermore, the various accounts are shaped by which explicandum they set out to explicate. How 'meaning', 'truth', and 'intention' shape the various conceptualizations depend on the scope of the different accounts as well as the explicanda. For instance, Fred Dretske (1981, 1983, 1988, 2008) develops an account of semantic information on the basis of an information-theoretic notion of information as explicandum. The purpose is to naturalize intentionality, explain the rise of meaning, offer an informational account of knowledge, and explain behavior. Therefore, Dretske's (1981) notion of information, as signals about events, is an objective, true, agent-independent commodity, where 'meaning' is not inherent but a subsequent ascription performed by agents. Due to the truth-requirement for information, the notions of misinformation and disinformation defined by falsity are not kinds of information (Dretske, 1981, 1983).

Floridi's (2004b, 2005a, 2005b, 2007, 2011b) notion of semantic information is developed within a conception of the universe as the Infosphere, where people as inforgs live their onlives in the middle of the information revolution. Floridi's notion of truthful semantic information is at the center of this world view and the overall project is to explain everything in informational terms. Therefore, information is conceived as objective and inherently truthful – thus, misinformation and disinformation, defined by falsity, are not kinds of information. The scope of Scarantino & Piccinini's (2010) account of natural and nonnatural information is to develop a conceptualization of information which reflects how the notion is used and understood in cognitive science and computer science. Their account is developed as an additional explication of the explicata provided by Dretske and Floridi with Grice's notions of natural and nonnatural meaning as further explicanda. The Gricean notion of nonnatural meaning as a basis for nonnatural information enables information to deal with conventions, intentions, and pragmatic aspects of 'meaning'. Scarantino and Piccinini (2010) argue that in order to fully capture these pragmatic features and in order to match how the notion is actually conceived of in cognitive science 'information' is alethically neutral.

Fox (1983) works with an explicandum of information which has to do with language, learning, and communication. Therefore, his explicatum is bound by language through a propositional account of information. Information is propositions expressed by sentences glued together by 'meaning' and therefore is agent-dependent. Fox (1983) argues for alethic neutrality for information and necessary falsity for misinformation, which is a kind of information. Fallis (2009, 2011, 2014, 2015) has his main emphasis on the notion of disinformation. The purpose is to offer a conceptual analysis which adequately account for all the different varieties of disinformation and which draws the distinctions between 'lies', 'disinformation', and 'deception'. In contrast to the explicanda of disinformation Fallis argues in favor of a true variety of disinformation based on false Gricean implicatures in line with the philosophical literature on lying, misleading, and deceiving (cf. Ch. 1).

Through the analyses and discussions of the interconnections of information, misinformation, and disinformation in the accounts put forth by Dretske, Floridi, Fox, Fallis, and Scarantino and Piccinini I find that 'truth' is secondary to 'meaning' and 'intention'. Although the debates about information often concern whether or not information requires truth, truth is neither a sufficient nor a necessary condition for information in the distinction between information, misinformation, and disinformation. That is, it is not a requirement for truth which distinguishes information from misinformation and disinformation. Misinformation and disinformation are not necessarily false. Fallis (2009, 2011, 2014, 2015) argues in favor of true disinformation and I extend his argument to include true misinformation as well. Therefore, truth in itself does not secure 'information' in the sense that truth does not necessarily prevent 'misleadingness'. The real distinguishing and definitional aspects are

intention/intentionality (i.e. the intention, intentionality, or un-intention with which some representational content is provided) and misleadingness/non-misleadingness as a pragmatic feature of 'meaning'. Thus, the truth-value of some semantic or representational content is secondary to the intention behind it and the misleadingness/non-misleadingness of it in order to determine whether the semantic/representational content is information, misinformation, or disinformation. However, a notion of information as non-misleading is not provided in the current accounts of information. Neither is a unified conceptualization of information, misinformation, and disinformation where all three notions are explored and analyzed to the same extent.

The collapse of the truth-false dichotomy for information vs. misinformation and disinformation and the recognition that misleadingness and non-misleadingness along with intention and intentionality are the real distinguishing features of the three notions are revealed through the analyses of the interconnections of information, misinformation, and disinformation. It is only apparent when all three notions are taken into account that truth and falsity do not distinguish the notions satisfactorily.

When the various explicata are discussed in relation to one another it shows that they cannot be combined in a unified conceptualization of information, misinformation, and disinformation as they stand. As the explicata of information, misinformation, and disinformation, respectively, are not defined in relation to one another, i.e. in accordance to their interconnections, it is not possible to pick three explicata (one for information, one for misinformation, and one for disinformation) and combine them in a unified conceptualization. However, it is such a unified conceptualization which is needed in order to automatically detect information, misinformation, and disinformation.

I therefore develop a unified conceptualization of information, misinformation, and disinformation which consists of three levels, four different notions of information, one notion of misinformation, and one notion of disinformation.

The first level consists of one notion:

- "Information" as an overall term

which refers to all the different explicanda of information. It is the conglomerate of all the 'ordinary notions' of information; the neutral mass noun; the concept which is used when a specific understanding of information is not specified.

The second level consists of two different notions of information:

- Natural information and
- Nonnatural information

based on the proposal by Scarantino and Piccinini (2010). Natural information is needed in order to naturalize intentionality whereas nonnatural information, which is roughly equivalent to representational content, is connected to conventions, language, and communication. Both natural and nonnatural information are alethically neutral and they are based on Grice's (1957) distinction between natural and nonnatural meaning. Thus, Grice's conception of 'meaning' explicitly shapes the conceptualization of information, misinformation, and disinformation and marks the difference between physical occurrences in the world and human constructs such as conventions and communication.

The third level consists of three different notions which are aspects of nonnatural information:

- Information as intentional non-misleading,
- Misinformation as unintended misleading, and
- Disinformation as intentional misleading.

'Misleadingness' and 'non-misleadingness', as pragmatic features of 'meaning', are partly determined by Gricean implicatures. An implicatum counts as part of the content of an utterance (e.g. oral, written, gesture, etc.) although it is not part of what is literally said. An utterance which is literally true can have a false implicatum and thereby generate a false implicature, whereby it is potentially misleading and can be true misinformation or true disinformation (dependent on intention/intentionality). An utterance which is literally false can have a true implicatum and thereby generate a true implicature (e.g. sarcasm and irony), whereby it is not misleading and thus is intentionally non-misleading information. Therefore, intentionally non-misleading information, misinformation (i.e. unintended misleading), and disinformation (i.e. intentional misleading) are defined as alethically neutral in order to capture all the different varieties of each notion.

Furthermore, Grice's (1967) Cooperative Principle and its maxims shape the unified conceptualization in the sense that they can be used as a heuristic for the determination of whether some piece of nonnatural information – i.e. representational content – is intentional non-misleading information, misinformation, or disinformation.

Perspectives

One thing is the desire to automatically detect misinformation and disinformation, as well as information in online social network structures in order to prevent potential dangers. Another thing is the potential need to use information, misinformation, and disinformation to protect one's privacy within these online social network structures. It is well known that social networks such as Facebook log all the data they can about their users in order to

generate profiles which can be used for targeted advertising. However, the data which is collected about specific people through the internet can be used for other purposes than advertising. For instance, Frank Pasquale, in his recent book *The Black Box Society* (2015), points to American banks' use of customers' Facebook profiles in order to assign their credit scores. The credit scores of potential customers were partly determined by the assigned credit scores of the customer's Facebook-friends.

Disinformation can be used as a coping-strategy in connection to protection of one's privacy. At a panel discussion following the pre-screening of the documentary *Facebookistan* at the University of Copenhagen, it was pointed out that some people deliberately try to mislead the algorithms running behind Facebook in order to disrupt the profile the algorithms generate. In this way, some people use personal disinformation in order to protect their personal information. For instance, people can 'like' posts, pages, and articles which they do not actually like and they can randomly visit pages and profiles. Furthermore, they can post personal disinformation – that is disinformation about themselves – in order to protect their personal information.

The unified conceptualization of information, misinformation, and disinformation developed in this thesis along with the analyses and discussions that precede it might provide a useful theoretical foundation for the further investigation of personal disinformation, personal misinformation, and personal information in connection to privacy issues.

References

- Adams, F. (2015). Personal Communication. Conversation, Apr 24th 2015.
- Adams, F. and de Moraes, J. A. (forthcoming). Is There a Philosophy of Information? *Topoi*. DOI 10.1007/s11245-014-9252-9. Springer, published online 2014.
- Adler, J. (1997). Lying, Deceiving, or Falsely Implicating. *Journal of Philosophy*, 94, pp. 435-452.
- Adriaans, P. (2012). *Information*. Stanford Encyclopedia of Philosophy. Zalta, E. N. (ed.), First published Fri Oct 26, 2012. (Fall 2013 Edition), retrieved at <http://plato.stanford.edu/archives/fall2013/entries/information/>
- Adriaans, P. and van Benthem, J. (2008). Introduction: Information is what Information does. In: *Philosophy of Information*, 2008, Adriaans, P. and van Benthem, J. (eds.), Elsevier B. V. pp. 2-26.
- Adriaans, P. and van Benthem, J. Eds. (2008). *Philosophy of Information*. Elsevier B. V.
- Austin, J. L. (1962). *How to Do Things with Words*. Second edition (1975), Urmson, J. O. and Sbisà, M. (eds.). Harvard university Press, Cambridge, Massachusets.
- BBC News, Technology (2014). *Lie detector on the way to test social media rumours*. Published Feb 19, 2014. Retrieved at <http://www.bbc.co.uk/news/technology-26263510?print=true>
- Bach, K. (2006). The top 10 misconceptions about implicature. In: *Drawing the Boundaries of Meaning: Neo-Gricean Studies in Pragmatics and Semantics in Honor of Laurence R. Horn*. Birner, B. and Ward, G. (eds.). John Benjamins Publishing Company, pp. 21-30.

- Blair, D. C. (1992). Information Retrieval and the Philosophy of Language. *The Computer Journal*, 35(3), pp. 200-207.
- Blair, D. C. (2003). Information Retrieval and the Philosophy of Language. *Annual Review of Information Science and Technology*, 37, pp. 3-50.
- Bontcheva, K. (2014). PHEME overview presentation. Retrieved at <http://www.pheme.eu/wp-content/uploads/2014/04/PHEME-overview-presentation.pdf>
- Brown, J. and Cappelen, H. Eds. (2011). *Assertion. New Philosophical Essays*. Oxford University Press.
- Budd, J. M. (2011). Meaning, truth, and information: prolegomena to a theory. *Journal of Documentation*, 67(1), pp. 56-74.
- Capurro, R. and Hjørland, B. (2003). The Concept of Information. *Annual Review of Information Science and Technology*, 37, pp. 343-411.
- Carnap, R. (1950). On explication. In: *Logical Foundations of Probability*. Routledge and Kegan Paul, Ltd. pp. 1-18.
- Carnap, R. and Bar-Hillel, Y. (1952). *An Outline of a Theory of Semantic Information*. Technical Report No. 247, Research Laboratory of Electronics, MIT.
- Chalmers, D. J. (2011). Verbal Disputes. *Philosophical Review*, 120(4), pp. 515-566.
- Chapman, S. (2005). *Paul Grice, Philosopher and Linguist*. Palgrave Macmillan.
- Collin, F. and Guldmann, F. (2005). *Meaning, Use and Truth. Introducing the Philosophy of Language*. Ashgate.
- Cornelius, I. (1996). *Meaning and Method in Information Studies*. Ablex Publishing Corporation.
- Cornelius, I. (2002). Theorizing Information for Information Science. *Annual Review of Information Science and Technology*, 36, pp. 393-425.
- David, M. (2002) *The Correspondence Theory of Truth*. Stanford Encyclopedia of Philosophy. Zalta, E. N. (ed.), First published Fri May 10, 2002; substantive revision Thu Jul 2, 2009. (Fall 2015 Edition), retrieved at <http://plato.stanford.edu/archives/fall2015/entries/truth-correspondence/>

- Davidson, D. (2001). Essay 5. Intending (1978). In: *Essays on Actions and Events*, 2001, Oxford Scholarship Online. Retrieved at www.oxfordscholarship.com
- Davis, W. (2005). *Implicature*. Stanford Encyclopedia of Philosophy. Zalta, E. N. (ed.), First published Fri May 6, 2005; substantive revision Tue Jun 24, 2014. (Fall 2014 Edition), retrieved at <http://plato.stanford.edu/archives/fall2014/entries/implicature/>
- Day, R. E. (2001). *The Modern Invention of Information*. Southern Illinois University Press.
- Derczynski, L. and Bontcheva, K. (2014a). *Spatio-temporal grounding of claims made on the web, in PHEME*. Proceedings from isa-10: 10th joint ACC-ISO SIGSEM Workshop on Interoperable Semantic Annotation, May 26, 2014, Reykjavik, Iceland. Retrieved at http://sigsem.uvt.nl/isa10/ISA-10_proceedings.pdf#page=73
- Derczynski, L. and Bontcheva, K. (2014b). *PHEME: Veracity in Digital Social Networks*. Proceedings, UMAP Project Synergy (UMAP Pros) Workshop, July 7, 2014. Retrieved at <http://derczynski.com/sheffield/papers/synergy-workshop.pdf>
<https://umapprox.wordpress.com/program/>
- Dretske, F. (1981). *Knowledge and the Flow of Information*. MIT Press, first paperback edition 1983.
- Dretske, F. (1983). *Précis of Knowledge and the Flow of Information. The Behavioral and Brain Sciences*, 6, pp. 55-63.
- Dretske, F. (1986). Misrepresentation. In: *Readings in philosophy and cognitive science*. Goldman, A. (ed.), MIT Press, 1993. "A Bradford Book". pp. 297-314.
- Dretske, F. (1988). Ch. 3 Representational Systems. In: *Explaining Behavior*. MIT Press. "A Bradford Book".
- Dretske, F. (2008). Epistemology and Information. In: *Philosophy of Information*, 2008, Adriaans, P. and van Benthem, J. (eds.), Elsevier B. V. pp. 29-47.
- Fallis, D. (2009). *A Conceptual Analysis of Disinformation*. Preprint from iConference 2009. Retrieved at https://www.ideals.illinois.edu/bitstream/handle/2142/15205/fallis_disinfo1.pdf?sequence=2 through Google Scholar.
- Fallis, D. (2010). Lying and Deception. *Philosophers' Imprint*, 10(11).
- Fallis, D. (2011). Floridi on Disinformation. *Etica & Politica*, 13(2), pp. 201-214.

- Fallis, D. (2014). The Varieties of Disinformation. In: *The Philosophy of Information Quality*, 2014, Floridi, L. and Illari, P. (eds.), Synthese Library 358, Springer. pp. 135-161.
- Fallis, D. (2015). What is Disinformation? *Library Trends*, 63(3), pp. 401-426.
- Fallis, D. (2015). Personal communication. E-mail message to author, Aug 14th 2015.
- Fetzer, J. H. (2004a). Information: Does it Have to be True? *Minds and Machines*, 14, pp. 223-229.
- Fetzer, J. H. (2004b). Disinformation: The Use of False Information. *Minds and Machines*, 14, pp. 231-240.
- Floridi, L. (1999). Information ethics: On the philosophical foundation of computer ethics. *Ethics and Information Technology*, 1, pp. 37-56.
- Floridi, L. (2002a). On defining library and information science as applied philosophy of information. *Social Epistemology: A Journal of Knowledge, Culture and Policy*, 16(1), pp. 37-49.
- Floridi, L. (2002b). What is the Philosophy of Information? *Metaphilosophy*, 33(1/2), 123-145.
- Floridi, L. (2003). Chapter 4. Information. In: *The Blackwell Guide to the Philosophy of Computing and Information*, Floridi, L. (ed.). Blackwell.
- Floridi, L. (2004a). LIS as Applied Philosophy of Information: A Reappraisal. *Library Trends*, 52(3), pp. 658-665.
- Floridi, L. (2004b). Outline of a Theory of Strongly Semantic Information. *Minds and Machines*, 14, pp. 197-222.
- Floridi, L. (2005a). Is Semantic Information Meaningful Data? *Philosophy and Phenomological Research*, LXX(2), pp. 351-370.
- Floridi, L. (2005b). *Semantic Conceptions of Information*. Stanford Encyclopedia of Philosophy. Zalta, E. N. (ed.), First published Wed Oct 5, 2005; substantive revision Mon Feb 4, 2013. (Spring 2013 Edition), retrieved at <http://plato.stanford.edu/archives/spr2013/entries/information-semantic/>
- Floridi, L. (2007). In Defence of the Veridical Nature of Semantic Information. *European Journal of Analytic Philosophy* (EUJAP), 3(1), pp. 31-41.
- Floridi, L. (2008). Trends in the Philosophy of Information. In: *Philosophy of Information*, 2008, Adriaans, P. and van Benthem, J. (eds.), Elsevier B. V. pp. 113-131.

- Floridi, L. (2011a). A Defence of Constructionism: Philosophy as Conceptual Engineering. *Metaphilosophy*, 42(3), pp. 282-304.
- Floridi, L. (2011b). *The Philosophy of Information*. Oxford Scholarship Online. Retrieved at www.oxfordscholarship.com
- Fox, C. J. (1983). *Information and Misinformation. An Investigation of the Notions of Information, Misinformation, Informing, and Misinforming*. Greenwood Press.
- Frege, G. (1892). On Sense and Reference. In: *Basic Topics in the Philosophy of Language*. Harnish, R. M. (ed.), 1994, Prentice-Hall, Inc. pp. 142-160.
- Furner, J. (2004). Information Studies Without Information. *Library Trends* 52(3), pp. 427-446.
- Furner, J. (2010). Philosophy and Information Studies. *Annual Review of Information Science and Technology*, 44(1), pp. 154-200.
- Furner, J. (2014). Information Without Information Studies. In: *Theories of Information, Communication and Knowledge*, Studies in History and Philosophy of Science 34, 2014, Ibekwe-SanJuan, F. and Dousa, T. M. (eds.). Springer Science+Business Media B. V. pp. 143-179.
- Godfrey-Smith, p. (1989). Misinformation. *Canadian Journal of Philosophy*, 19(4), pp. 533-550.
- Grice, H. P. (1957). Meaning. In: *Studies in the Way of Words*, 1989. First Harvard University Press Paperback Edition, 1991. pp. 213-223.
- Grice, H. P. (1958). Postwar Oxford Philosophy. In: *Studies in the Way of Words*, 1989. First Harvard University Press Paperback Edition, 1991. pp. 171-180.
- Grice, H. P. (1967). Logic and Conversation. In: *Studies in the Way of Words*, 1989. First Harvard University Press Paperback Edition, 1991. pp. 22-40.
- Grice, H. P. (1971). Intention and Uncertainty. *Proceedings of the British Academy, London*, 57, pp. 263-279.
- Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press. First Paperback Edition 1991.
- Harnish, R. M. (ed.) (1994). *Basic Topics in the Philosophy of Language*. Prentice Hall.

- Hintikka, J. (2007). Epistemology without Knowledge and without Belief. In: *Socratic Epistemology. Explorations of Knowledge-Seeking by Questioning*. Cambridge University Press. pp. 11-37.
- Jacob, P. (2003). *Intentionality*. Stanford Encyclopedia of Philosophy. Zalta, E. N. (ed.), First published Thu Aug 7, 2003; substantive revision Wed Oct 15, 2014. (Winter 2014 Edition), retrieved at <http://plato.stanford.edu/archives/win2014/entries/intentionality/>
- Karlova, N. A. and Fisher, K. E. (2013). A social diffusion model of misinformation and disinformation for understanding human information behavior. *Information Research*, 18(1).
- Karlova, N. A. and Lee, J. H. (2011). Notes from the Underground City of Disinformation: A conceptual Investigation. *ASIST*.
- Korta, K. and Perry, P. (2006). *Pragmatics*. Stanford Encyclopedia of Philosophy. Zalta, E. N. (ed.), First published Tue Nov 28, 2006; substantive revision Tue Apr 2, 2015. (Winter 2015 Edition), retrieved at <http://plato.stanford.edu/archives/win2015/entries/pragmatics/>
- Kumar, K. P. K. and Geethakumari, G. (2014). Detecting misinformation in online social networks using cognitive psychology. *Human-centric Computing and Information Sciences*, 4(1).
- Künne, W. (2003). *Conceptions of Truth*. Oxford University Press. Through: Oxford Scholarship Online. Retrieved at www.oxfordscholarship.com
- Lenski, W. (2010). Information: A Conceptual Investigation. *Information*, 1, pp. 74-118.
- Leonard, R. A. (2015). Personal communication, conversation after the lecture *Language can solve – and prevent – crimes*, University of Copenhagen, Jun 8th 2015.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.
- Mahon, J. (2008). *The Definition of Lying and Deception*. Stanford Encyclopedia of Philosophy. Zalta, E. N. (ed.), First published Thu Feb 21, 2008. (Fall 2009 Edition), retrieved at <http://plato.stanford.edu/archives/fall2009/entries/lying-definition/>
- Mai, J.-E. (2013). The quality and qualities of information. *Journal of the American Society for Information Science and Technology*, 64(4), pp. 675-688.

- McGrath, M. (2005). *Propositions*. Stanford Encyclopedia of Philosophy. Zalta, E. N. (ed.), First published Mon Dec 19, 2005; substantive revision Wed Jun 20, 2012. (Spring 2014 Edition), retrieved at <http://plato.stanford.edu/archives/spr2014/entries/propositions/>
- Mele, A. R. (1992). *Springs of Action : Understanding Intentional Behavior*. Cary, NC, USA: Oxford University Press, USA. ProQuest ebrary. Web. 12 January 2015.
- Mele, A. R. (2010). Intention. In: *A Companion to the Philosophy of Action*, 2010, O'Connor, T. and Sandis, C. (eds.). Blackwell Publishing Ltd. pp. 108-113.
- Merriam-Webster (1991). *New Book of Word Histories*. Mish, F. C. (ed.) Merriam-Webster, US.
- Olsson, E. J. (2015). Gettier and the method of explication: a 60 year old solution to a 50 year old problem. *Philosophical Studies*, 172(1), pp. 57-72.
- Pasquale, F. (2015). *The Black Box Society. The Secret Algorithms That Control Money and Information*. Harvard University Press.
- PHEME (2014). *About PHEME*. Retrieved at www.pHEME.eu, 27.03.14.
- Robinson, L. and Bawden, D. (2014). Mind the Gap: Transitions Between Concepts of Information in Varied Domains. In: *Theories of Information, Communication and Knowledge*, Studies in History and Philosophy of Science 34, 2014, Ibekwe-SanJuan, F. and Dousa, T. M. (eds.). Springer Science+Business Media B. V. pp. 121-141.
- Scarantino, A. and Piccinini, G. (2010). Information Without Truth. *Metaphilosophy*, 41(3), pp. 313-330.
- Setiya, K. (2009). *Intention*. Stanford Encyclopedia of Philosophy. Zalta, E. N. (ed.), First published Mon Aug 31, 2009; substantive revision Mon Jan 20, 2014. (Summer 2015 Edition), retrieved at <http://plato.stanford.edu/archives/sum2015/entries/intention/>
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, pp. 379-423, 623-656, July, October.
- Sheffield News (2014). *EU project to build lie detector for social media*. Published Feb 18, 2014. Retrieved at www.sheffield.ac.uk/news/nr/lie-detector-social-media-sheffield-twitter-facebook-1.354715
- Simpson, J. (2013). The *Oxford English Dictionary* and Its Philosophy. *Philosophy and Technology*, 26, pp. 341-348.
- Skyrms, B. (2010). *Signals. Evolution, Learning, & Information*. Oxford University Press.

- Smid, J. (2015). Personal communication. E-mail message to author, Nov 17th 2015.
- Speaks, J. (2010). *Theories of Meaning*. Stanford Encyclopedia of Philosophy. Zalta, E. N. (ed.), First published Tue Jan 26, 2010, substantive revision Wed Apr 23, 2014. (Fall 2014 Edition), retrieved at <http://plato.stanford.edu/archives/fall2014/entries/meaning/>
- Stokke, A. (2013). Lying, Deceiving, and Misleading. *Philosophy Compass*, 8(4), pp. 348-359.
- Stoljar, D. and Damnjanovic, N. (1997). *The Deflationary Theory of Truth*. Stanford Encyclopedia of Philosophy. Zalta, E. N. (ed.), First published Thu Aug 28, 1997; substantive revision Mon Oct 4, 2010. (Fall 2014 Edition), retrieved at <http://plato.stanford.edu/archives/fall2014/entries/truth-deflationary/>
- Søe, S. O. (submitted). Intention and Misleading.
- Van der Veer Martens, B. (2015). An Illustrated Introduction to the Infosphere. *Library Trends*, 63(3), pp. 317-361.
- Webber, J. (2013). Liar! *Analysis*, 73(4), pp. 651-659.
- Wilson, P. (1977). *Public Knowledge, Private Ignorance. Toward a Library and Information Policy*. Greenwood Press.
- Wilson, G. and Shpall, S. (2002). *Action*. Stanford Encyclopedia of Philosophy. Zalta, E. N. (ed.), First published Mon Mar 18, 2002; substantive revision Wed Apr 4, 2012. (Summer 2012 edition), retrieved at <http://plato.stanford.edu/archives/sum2012/entries/action/>
- The Society for the Philosophy of Information. About PI. Retrieved at <http://www.socphilinfo.org/about-pi/about-pi>
- The Π Research Network (2013): *The Philosophy of Information – An Introduction*. Version 1.0, The Society for the Philosophy of Information. Retrieved at http://socphilinfo.org/sites/default/files/i2pi_2013.pdf

Online dictionaries, Information:

Dictionary.com: <http://dictionary.reference.com/browse/information?s=t>

Merriam-Webster, Learner's Dictionary:

<http://www.learnersdictionary.com/definition/information>

Oxford Dictionaries: <http://www.oxforddictionaries.com/definition/english/information>

Oxford English Dictionary:

<http://www.oed.com/view/Entry/95568?redirectedFrom=information#eid>

Vocabulary.com: <http://www.vocabulary.com/dictionary/information>

Online dictionaries, Misinformation:

Britannica, Merriam-Webster:

<http://www.britannica.com/bps/dictionary?query=misinformation>

Dictionary.com: <http://dictionary.reference.com/browse/misinformation>

Merriam-Webster, Learner's Dictionary:

<http://www.learnersdictionary.com/definition/misinformation>

Oxford Dictionaries: <http://www.oxforddictionaries.com/definition/english/misinformation>

Oxford English Dictionary: <http://www.oed.com/view/Entry/119699#eid36703738>

Vocabulary.com: <http://www.vocabulary.com/dictionary/misinformation>

Online dictionaries, Disinformation:

Britannica, Merriam-Webster:

<http://www.britannica.com/bps/dictionary?query=disinformation>

Merriam-Webster, Learner's Dictionary:

<http://www.learnersdictionary.com/definition/disinformation>

Dictionary.com: <http://dictionary.reference.com/browse/disinformation?s=t>

Oxford Dictionaries: <http://www.oxforddictionaries.com/definition/english/disinformation>

Oxford English Dictionary:

<http://www.oed.com/view/Entry/54579?redirectedFrom=disinformation#eid>

Vocabulary.com: <http://www.vocabulary.com/dictionary/disinformation>